PROCEEDINGS IGS 2025

Investigating Human Movements - Handwriting and Beyond

22nd Conference of the
International Graphonomics Society

Réjean Plamondon
Céline Rémi
Karina Lebel
Mickaël Begon
Frederic Fol Leymarie
Lama Séoud
Benjamin De Leener
Eva Alonso Ortiz
Editors

June 2-5, 2025, Montréal, Canada



COLLECTION PRESSES INTERNATIONALES POLYTECHNIQUE

Les Presses de l'Université de Montréal

Réjean Plamondon
Institut de génie biomédical, Département de génie électrique, Polytechnique Montréal, Canada
Céline Rémi
Laboratoire de Mathématiques Informatique et Applications, Université des Antilles, Guadeloupe, France
Karina Lebel
Centre de recherche sur le vieillissement, Département de génie électrique et de génie informatique, Université de Sherbrooke, Canada
Mickaël Begon
École de kinésiologie et des sciences de l'activité physique, Faculté de médecine, Université de Montréal, Canada
Frederic Fol Leymarie
Goldsmiths, School of Computing, University of London, United Kingdom
Lama Séoud
Institut de génie biomédical, Département de génie informatique et génie logiciel, Polytechnique Montréal, Canada
Benjamin De Leener
Institut de génie biomédical, Département de génie informatique et génie logiciel, Polytechnique Montréal, Canada
Eva Alonso Ortiz
Institut de génie biomédical, Département de génie électrique, Polytechnique Montréal, Canada
Editors

PROCEEDINGS IGS 2025

Investigating Human Movements - Handwriting and Beyond

22nd Conference of the
International Graphonomics Society

Montréal, Canada, June 2-5, 2025

COLLECTION PRESSES INTERNATIONALES POLYTECHNIQUE

Les Presses de l'Université de Montréal

To quote a paper from this proceedings:

Author(s) Initial(s). Surname(s), "Title of paper," in *Proc. IGS 2025: Investigating Human Movements – Handwriting and Beyond*, Plamondon, R., Rémi, C., Lebel, K., Begon, M., Fol Leymarie, F., Séoud, L., De Leener, B., Alonso-Ortiz, E., (Eds) Presented at the 22nd Conf. of the Int. Graphonomics Society, Polytechnique Montréal, Montréal, Canada, 2025, pp. 256.
[Online]. Available: https://pum.umontreal.ca/catalogue/proceedings_igs_2025/fichiers

Access the Proceedings' Web page here: https://pum.umontreal.ca/catalogue/proceedings_igs_2025
E-mail Les Presses de l'Université de Montréal at: pum@umontreal.ca

Québec
Books Tax Credit
Gestion SODEC

Funded by the Government of Canada
Canada

All rights reserved.
© Presses de l'Université de Montréal and the authors.
This book may not be duplicated in any way without the express written consent of the publisher.

Legal deposit: 1st quarter 2026
Bibliothèque et Archives nationales du Québec
Library and Archives Canada

9 782760 655096
ISBN 978-2-7606-5509-6

22nd CONFERENCE OF THE

International Graphonomics Society

June 2-5, 2025
Polytechnique Montréal
Canada





Partners

IGS INTERNATIONAL GRAPHONOMICS SOCIETY

POLYTECHNIQUE MONTRÉAL
TECHNOLOGICAL UNIVERSITY

Direction de la recherche et de l'innovation
Département de génie électrique
Département de génie informatique et génie logiciel
Département de génie mécanique

Scribens

NeuroScribens

INSTITUT TRANSMEDTECH MONTRÉAL LIVING LAB

CENPARMI
Centre for Pattern Recognition and Machine Intelligence

NeuroScript

Investigating Human Movements | Handwriting and Beyond

Table of contents

Keynotes

Catalyzing Innovation in Healthcare Technologies: The TransMedTech's Living Lab Platform 2
Carl-Éric Aubin

Tablet Technology Compatible with Functional Magnetic Resonance Imaging: Exploring the Brain Activity Associated with Writing and Drawing Behaviour 4
Simon J. Graham

Movement Generation and Drawing in Robotics 5
Sylvain Calinon

Biomarkers

Lognormality-Based Aggregation to Assist Kindergarten Pupil's Graphomotor Skills Analysis 10
Céline Rémi, Sébastien Regis, Jean Vaillant, Réjean Plamondon

Handwriting Analysis for the Detection of Lewy Body Dementia: Methods, Challenges, and Perspectives 15
Tiziana D'Alessandro, Claudio De Stefano, Francesco Fontanella, Emanuele Nardone, Olga Pustovalova, Alessandra Scotto di Freca

Handwriting Measures as Predictors of Executive Functioning, Time Organization and Self-Estimation of Students with and without Specific Learning Disabilities 20
Sara Rosenblum, Kineret Sharfi

Leveraging Grammatical Evolution for Interpretability in Computer-Assisted Diagnosis of Alzheimer's Disease 25
Amos Battista, Antonio Della Cioppa, Angelo Marcelli

Multimodal Handwriting Analysis in Alzheimer's Disease Detection through Bayesian Networks 30
Emanuele Nardone, Tiziana D'Alessandro, Claudio De Stefano, Francesco Fontanella, Alessandra Scotto di Freca

Student Competition

Biceps Curl Exercise Analysis with Wearable Smartwatch Sensors 35
Cristian Rodriguez, David Espino, Cristina Jimenez-Marcos, Moises Diaz, Cristina Carmona-Duarte, Miguel A. Ferrer

The Effect of Keyboard Size and Anthropometry on Pianists' Finger Kinematics 39
Craig Turner, Robin Mailly, Justine Pelletier, Mickaël Begon, Felipe Verdugo

Enhancement of IntuiSketch ITS with Teacher's Planned Sequence to Improve Learning by Drawing 43
Islam Barchouch, Swan Serre, Eric Anquetil, Nathalie Girard

Hypothesis-Aware Ductus Generation for Greek Papyri: A Data-Driven Approach 47
Dinh Dung Van, Giuseppe De Gregorio, Rodrigo Cerqueira Gonzalez Pena

A Dual Backbone Framework for Alzheimer's Detection through Handwriting Multiple Sources 52
Cesare Davide Pace, Emanuele Nardone, Tiziana D'Alessandro

Alzheimer's Disease Screening with Limited and Noisy On-line Pentagon Drawings 57
Nina Hosseini-Kivanani, Elena Salobrar-García, Moises Diaz, Miguel A. Ferrer, Luis A. Leiva

Special session: KNM2 & More for Potential Collaborations

Lognormal Function as a Primitive for Sinusoidal Waveforms: Case of the Velocity Profile in Oscillatory Human Movement 63
Moussa Djioua

The Pen-Stroke Test to Identify Suboptimal Neurodevelopment in Children Born Prematurely 68
Thuy Mai Luu, Guillaume Séguin de Broin, Mi-Suk Kang-Dufour, Victoria Côté-Piché, Léanne Brabant, Marie-Michèle Gagnon, Mélanie Gagnon, Catherine Bernard, Véronique Belval, Marie-Noëlle Simard, Réjean Plamondon, Mathieu Dehaes

KNM2: a Collaborative Research Platform for Lifelong Lognormality Monitoring 73
Réjean Plamondon, Moussa Djioua, Lisa Jithame, Lina Sehad, Rostand Marlone Njomo Njampou

Posters

A Benchmark Dataset for Handwritten Document Layout Analysis in the Wild 79
Marina Gardella

A Survey on Skeletonization of Character Images 84
Yunqi Xu, Ching Y. Suen

Age Estimation from Handwriting Images with Regression Models 88
Lior Abergel, Maor Merling, Marina Litvak, Irina Rabaev

Benchmarking Facial Landmarks Estimation: Evaluating Popular Algorithms Using FAIRSET, a Balanced Landmark Dataset 93
Ian-Mathieu Joly, Nikola Zelovic, Eléonor Riesco, Karina Lebel

Improving Personality Prediction in Handwriting Analysis through Optimal Input Granularity 97
Yan Xu, Yufang Tang, Ching Y. Suen

Comparing Character Shape between Two Japanese Handwriting Databases of Forty-Years' Difference 102
Yoshinori Akao

Development of an Automated Tool for the Extraction of Muscle Synergies from Electromyographic Data 106
Flore Le Blanc, Simon Brière, Daniel Bourbonnais, Camille Martin, Sylvie Lafrenaye, Guillaume Léonard

TagTrack: A Camera-Based System for Movement Tracking with Applications in Resistance Training 111
Sara Houde, Camille Martin, Eleonor Riesco, Renaud Tremblay, Karina Lebel

Temporal Memory Index: A Novel Feature for Handwriting Analysis in Developmental Dysgraphia 115
Jan Mucha, Jiri Mekyska, Katarina Zvoncakova

Neuroscience & Applications

An Examination of Handwriting Inclination and Size as an Indicator of Emotional States: Separating Fact from Pseudoscience. . . . 120
Marcos Faundez-Zanuy, Anna Esposito, Gennaro Cordasco

Early Detection and Classification of Parkinsonism: Leveraging the Lognormal Model to Aid Clinical Assessment . . . 125
Karina Lebel, Cristina Carmona-Duarte, Pierre Blanchet, Vanessa Bachir, Guillaume Seguin de Broin, Réjean Plamondon

Deep Temporal Modeling for Automated Differentiation of Parkinson's Disease and Progressive Supranuclear Palsy Based on Gait Analysis . . . 130
Cesare Davide Pace, Maria Gaglione, Alessandro Marco De Nunzio, Claudio De Stefano, Francesco Fontanella, Maria Francesca De Pandis, Mario Molinara

Modeling Action Potentials with Sigma-Lognormal Equations . . . 133
Réjean Plamondon, Moussa Djioua, Christian O'Reilly

Using the Kinematic Theory to Infer a New Optimal Control Cost Function . . . 138
Najoua Assila, Ben Braithwaite, Mickaël Begon, Réjean Plamondon

Handwriting

High-Efficient Handwritten Text Recognition: A Lightweight CNN-Transformer with Self-Supervised Deconvolutional Decoding . . . 144
Baha Edine Harrath, Rayane Bencharef, Mohammed Hamden, Mohamed Cheriet

Investigating the Effects of Multisensory Integration in Writing and Memorizing New Graphic Patterns . . . 149
Yoan Segura, Florence Bara, Julien Tardieu, Pierre-Vincent Paubel, Jérémy Danna

Influence of Sampling Rate on Handwriting in Children with Developmental Dysgraphia . . . 153
Jan Mucha, Jiri Mekyska, Katarina Zvoncakova, Adam Klocek

Handwritten Character Representation and Generation Based on Bézier Curves . . . 157
Aleksandra Dundic, Andreas Fischer, Anna Scius-Bertrand

Write on Paper and Get the Online Digital Trace: A New Era for Handwriting . . . 161
Florent Imbert, Yann Soullard, Eric Anquetil, Tanja Harbaum, Alexey Serdyuk, Fabian Kreß, Tim Hamann, Peter Kämpf

Evaluation of the Precision of Markerless Motion Capture for Pediatrics Upper-Limb Kinematics . . . 166
Alexandre Naaim, Julie Rozaire, Lucas Mercier, Sonia Duprey, Mickaël Begon

Robotic & Arts

Drawing, Mind, and Movement: Possible Directions for Technological Exploration . . . 171
Seymour, Simmons III

Predicting Differential Aesthetics in Offline Handwritten Signatures . . . 176
Cristian Rodriguez, Moises Diaz, Miguel A. Ferrer

Differentiable Rasterization of Minimum-Time Sigma-Lognormal Trajectories . . . 180
Daniel Berio, Sylvain Calinon, Réjean Plamondon, Frederic Fol Leymarie

Articulation-Based Retrieval of Stroke Gestures 185
Inês Cardoso Oliveira, Nuwan T. Attygalle, Réjean Plamondon, Luis A. Leiva

Analyzing Angular Velocity Profiles of UR5e Robotic Arm Joints During Signing 190
Moises Diaz, Miguel A. Ferrer, José J. Quintana, Adam Wolniakowski, Roman Trochimczuk, Konstantsin Miatliuk

Human Movement as a State Transition 194
Moussa Djioua

Forensic & Security

Analysis of Inclinations in Genuine Signatures and Skilled Forgeries 200
Marcos Faundez-Zanuy

Quasi-Synthetic Riemannian Data Generation for Writer-Independent Offline Signature Verification 205
Elias N. Zois, Moises Diaz, Salem Said, Miguel A. Ferrer

Associating Velocity Profiles to Ink Traces for Forensic Handwriting Examination 209
Rosanna Coccaro, Enrico Ferrentino, Antonio Parziale, Angelo Marcelli

Evaluating Word Embedding Models for Authorship Attribution 213
Hiba Fallatah, Ching Y. Suen

Towards a Gold Standard for AI Facial Landmarks Estimation: Constructing FAIRSET, a Balanced and Inclusive Landmark Database 218
Nikola Zelovic, Ian-Mathieu Joly, Eléonor Riesco, Karina Lebel

The AI Evolution in Forensic Handwriting Examination: The Road toward Implementation 222
Elisa van den Heuvel, Samiah Ibrahim

Graphonomics for Education

Comparative Study of Shape and Velocity Reproduction in Neurodevelopmental Disorders 227
Pablo Santana-Susilla, Antonio Parziale, M. Sol Fortea, Angelo Marcelli, Cristina Carmona-Duarte

Learning and Transfer of Graphomotor Skills in Typically Developing 7 to 10 Years Old Children and Adults 231
Laureen Josseron, Jérôme Clerc, Caroline Jolly

The Impact of Speed and Scale on Spiral Drawing in Children with Developmental Dysgraphia 236
Jiri Mekyska, Jan Mucha, Katarina Zvoncakova, Zoltan Galaz, Marcos Faundez-Zanuy

Author Index

Foreword

On October 19, 2023, in Evora, Portugal, the IGS board gave me the mandate to organize the 22nd IGS conference in Montréal. I accepted, promising that I would do my best to extend its horizons. After a few months of planning, I came with a theme: "Investigating Human Movements: Handwriting and Beyond."

I then reached out to colleagues from Polytechnique Montréal to join me in the local organizing committee (Lama Séoud, Benjamin De Leener and Eva Alonso Ortiz), and signed a contract with Conferium (https://conferium.ca/en), where we benefitted from the professional expertise of Lisa Nolet, Alexis Levasseur, and Pierre Bolduc.

Regarding the scientific content, I invited four collaborators from various disciplines to act as program chairs (Céline Rémi, Frederic Fol Leymarie, Karina Lebel, and Mickaël Begon). The whole organization has been done in coordination and in collaboration with the IGS board members.

A rigorous review process led by the program chairs has resulted in the selection of 50 papers. The presentation format for each selected paper was determined by the presence of at least two positive reviews from two different members of the international independent review panel. These presentations plus three Keynotes Speeches from Carl-Éric Aubin, Simon Graham, and Sylvain Calinon and eight innovative demonstrators and artistic performances were scheduled all along the three-days of the single-track conference program.

Overall, 80 researchers from 15 countries, participated in the conference that was held at Pavillon Lassonde, Polytechnique Montréal. Their presentations covered 12 topics dealing with methods for capturing, processing and investigating Human Movements, including recent applications in healthcare, education, sports, arts, forensic science, cybersecurity, anthropomorphic robots and artificial intelligence.

This event also owed its success to the financial support of our sponsors: the Polytechnique Montreal Research and Innovation Directorate and the Departments of Electrical, Computer & Software, and Mechanical Engineering (gold partners), as well as the TransMedTech Institute, NeuroScribens Inc., NeuroScript Inc., and CENPARMI (bronze partners).

These proceedings present the extended abstracts of all the papers and posters that were presented during the conference. Some of their authors will be invited to submit a full paper to three Journal Special Issues (Human Movement Science, Arend van Gemmert contact editor; Pattern Recognition Letters, Claudio di Stefano contact editor; and Journal of Forensic Documents Examination, Angelo Marcelli contact editor).

The order of the abstract presentation follows the conference program.

As recently reported in an email to the board from Professor Claudio di Stefano, the president of IGS: "The IGS2025 conference was undoubtedly a success both for the number and the quality of the papers presented, and for the perfect organization: the invited speakers were high-profile and engaging in their presentations, as were all the sessions where the chairmen stimulated an interesting debate with the active participation of all the congress participants, always present throughout the conference. The poster sessions were also very interesting, with the now-consolidated technique of a short oral presentation of all the papers to stimulate interest in visiting the poster and speaking directly with the authors. The social events were carefully

curated, and the social dinner was splendid with an extraordinary musical show."

I truly thank all my collaborators for the extraordinary work done and all the researchers who participated to the 22nd IGS conference at Polytechnique Montréal, making this meeting a scientific and social experience that we will remember for a long time. Sincerely,

Réjean Plamondon, Professor
P.Eng., M.Sc.A., Ph.D.
IGS2025 GENERAL CHAIR
NIAS, IAPR and IEEE Fellow
Life Fellow IEEE
Head of Scribens Laboratory
Department of Electrical Engineering
Polytechnique Montréal

Committees

Local Organizing Committee
Polytechnique Montréal, Canada

Réjean Plamondon, General Chair	Département de génie électrique
Eva Alonso Ortiz	Département de génie électrique
Benjamin De Leener	Département de génie informatique et génie logiciel
Lama Séoud	Département de génie informatique et génie logiciel
Amal Bennani (Administration)	Département de génie électrique
Laurène Cicéri (Finance)	Département de génie électrique
Neïla Kaou (Advisor)	Conseillère R&D, Institut TransMedTech

International Steering Committee
IGS board

Claudio De Stefano	Università di Cassino e del Lazio Meridionale, Italy
Angelo Marcelli	Università degli Studi di Salerno, Italy
Réjean Plamondon	Polytechnique Montréal, Canada
Hans-Leo Teulings	NeuroScript LLC, USA
Arend Van Gemmert	Louisiana State University, USA
Céline Rémi	Université des Antilles, Guadeloupe, France
Andreas Fischer	University of Applied Sciences and Arts Western Switzerland (HES-SO), Switzerland
Ana Rita Matias	Comprehensive Health and Research Center, Universidade de Évora, Portugal
Erika Griechisch	Cursor Insight, Hungary
Sara Rosenblum	University of Haifa, Israel
Heidi Harralson	East Tennessee State University, USA
Masaki Nakagawa	Tokyo University of Agriculture and Technology, Japan
Moises Diaz	Universidad de Las Palmas de Gran Canaria, Spain

Program Chairs

Céline Rémi General Program Chair, Growth and Learning	Laboratoire de Mathématiques Informatique et Applications (LAMIA), Université des Antilles, Guadeloupe, France

Mickaël Begon Program co-Chair, Performance and Rehabilitation	École de kinésiologie et des sciences de l'activité physique, Faculté de médecine, Université de Montréal, Canada
Karina Lebel Program co-Chair, Aging and Healthcare	Centre de recherche sur le vieillissement (CdRV), Département de génie électrique et de génie informatique, Université de Sherbrooke, Canada
Frederic Fol Leymarie Program co-Chair, Art, Robotics and Forensics	Goldsmiths, School of Computing, University of London, United Kingdom

Program Committee

Yoshinori Akao	National Research Institute of Police Science, Japan
Ali Reza Alaei	Southern Cross University, Australia
Eric Anquetil	INSA Rennes / IRISA Laboratory, France
François Bailly	Institut national de recherche en sciences et technologies du numérique (INRIA), Montpellier, France
Mickael Begon	Université de Montréal, Canada
Daniel Berio	Goldsmiths, University of London, United Kingdom
Hala Bezine	FSS-University of Sfax, Tunisie
Leslie Bishko	Emily Carr University of Art + Design, Vancouver, Canada
Nathalie Bonneton Botté	Université de Bretagne Occidentale, France
Sylvain Calinon	Idiap Research Institute, Switzerland
Cristina Carmona-Duarte	Universidad de Las Palmas de Gran Canaria, Spain
Yury Chernov	Institute for Handwriting Sciences, Switzerland
Marion Cossin	Center for Research, Innovation and Transfer in Circus Arts, Montréal, Canada
Tiziana D'Alessandro	Università di Cassino e del Lazio Meridionale, Italy
Jérémy Danna	Laboratoire CLLE, Université Toulouse - Jean Jaurès, France
Giuseppe De Gregorio	University of Basel, Switzerland
Olivier Deussen	University of Konstanz, Germany
Nicolò Di Toma	Training & development specialist, Treviso, Veneto, Italia
Moises Diaz	Universidad de Las Palmas de Gran Canaria, Spain
Moussa Djioua	Polytechnique Montréal, Canada
Peter Drotar	Department of Computers and Informatics, Technical University of Kosice, Slovakia
Jagoda Dzida	Adam Mickiewicz University, Poznan, Poland
Jose Echevarria	Adobe Research, USA

Mounîm A. El-Yacoubi	Institut Polytechnique de Paris, France
Marcos Faundez-Zanuy	Tecnocampus, Universitat Pompeu Fabra, Spain
Miguel A. Ferrer	Universidad de Las Palmas de Gran Canaria, Spain
Andreas Fischer	University of Applied Sciences and Arts Western Switzerland (HES-SO), Switzerland
Frederic Fol Leymarie	Goldsmiths, University of London, United Kingdom
Francesco Fontanella	Università di Cassino e del Lazio Meridionale, Italy
Sonia Garcia-Salicetti	Institut Polytechnique de Paris/Télécom SudParis, France
Nathalie Girard	SHADoc, CNRS - IRISA – Université de Rennes, France
Marta Gomez-Barrero	Universität der Bundeswehr München, Germany
Etienne Goubault	Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST), Canada
Erika Griechisch	University of Szeged, Hungary
Caroline Jolly	LPNC, CNRS, Université Grenoble Alpes, Grenoble, France
Dahlia Kairy	Université de Montréal, Canada
Craig Kaplan	School of Computer Science, University of Waterloo, Canada
Karina Lebel	Département de génie électrique et de génie informatique, Université de Sherbrooke, Canada
Luis Leiva	University of Luxembourg, Luxembourg
Angelo Marcelli	DIEM - Università di Salerno, Italy
Ana Rita Matias	Comprehensive Health and Research Center, Universidade de Évora, Portugal
Mario Molinara	Università di Cassino e del Lazio Meridionale, Italy
Linton Mohammed	Forensic Science Consultants, Inc., USA
Ján Mucha	Brno University of Technology, Czech Republic
Alexandre Naaim	Laboratoire de Biomécanique et Mécanique des Chocs, Université Gustave Eiffel, Lyon, France
Masaki Nakagawa	Tokyo University of Agriculture and Technology, Japan
Adina Panchea	Département de génie électrique et informatique, Université de Sherbrooke, Canada
Theodoros Papatheodorou	Computational media arts department, Hong Kong University of Science and Technology, Guangzhou, China
Antonio Parziale	University of Salerno, Italy
Réjean Plamondon	Département de génie électrique, Polytechnique Montréal, Canada
Ioannis Pratikakis	Democritus University of Thrace (DUTH), Greece
Lionel Prevost	ESIEA, France
Sébastien Regis	French West Indies University, Guadeloupe, France

Céline Rémi	Université des Antilles, Guadeloupe, France
Sara Rosenblum	University of Haifa, Israel
Robert Sabourin	École de technologie supérieure, Canada
Anna Scius-Bertrand	Haute école spécialisée de Suisse occidentale de Fribourg (HES-SO// FR), Switzerland
Alessandra Scotto di Freca	Università di Cassino e del Lazio Meridionale, Italy
Seymour Simmons III	Winthrop University, USA
Michael Stroh	University of Konstanz, Germany
Jean Vaillant	Université des Antilles, Guadeloupe, France
Arend Van Gemmert	Louisiana State University, USA
Felipe Verdugo	École de kinésiologie et des sciences de l'activité physique, Faculté de médecine, Université de Montréal, Canada
Nicole Vincent	Université Paris Cité, France
Annie Vinter	Université Bourgogne Europe, France
Emily Will	Board of Forensic Document Examiners, USA





Session

Keynotes

Catalyzing Innovation in Healthcare Technologies: the TransMedTech's Living Lab Platform

Carl-Éric Aubin
Professor, O.C., O.Q., P.Eng., Ph.D., ScD (h.c.), Ch.O.M., FCAE
Chief Executive and Scientific Officer, TransMedTech Institute, Montréal, Québec, Canada
carl-eric.aubin@polymtl.ca

The health technology and innovation sector continues to grow rapidly worldwide and is central to the transformation of healthcare systems. Aging populations, rising healthcare costs, workforce shortages, and ongoing socio-political and environmental disruptions are compounding pressures on already strained health systems. These challenges have accelerated the need for faster, more inclusive, accessible, and more responsible innovation models. The post-COVID era has further underscored the importance of adopting agile, collaborative, impact-driven, and accessibility-conscious approaches that ensure equitable benefit across diverse user populations.

We are witnessing unprecedented momentum in the development of advanced medical technologies, digital health tools, and AI-powered solutions. Yet, this rapid growth unfolds within complex, highly regulated healthcare systems, often creating tension between the speed of innovation and the practicalities of clinical adoption. Simultaneously, societal expectations are evolving: innovation must now embody values of equity, inclusiveness, accessibility—including physical, digital, and cognitive dimensions—environmental sustainability, and long-term value creation that serves both individuals and communities.

Responding to these imperatives requires that innovation ecosystems evolve continuously toward models that are not only technologically robust, but also socially responsible, ecologically viable, and attuned to diverse user needs. Sustainable, accessible, interoperable, and value-based innovation must become the norm rather than the exception. Open innovation models—such as Living Labs—provide a powerful framework to achieve this transformation. The TransMedTech Institute exemplifies such a model. Operating as a Living Lab embedded within an ISO 13485-certified quality management system, TransMedTech mobilizes a diverse community of researchers, engineers, students, clinicians, patients, caregivers, industry leaders, health system stakeholders, and policy-makers. These actors collaborate as co-innovators throughout the entire lifecycle of health technologies—from research and validation to integration and implementation in real-world environments.

This collaborative and interdisciplinary ecosystem has catalyzed over 140 healthtech projects, supported through customized intermediation services, co-development platforms, regulatory strategy support, and clinical integration pathways. TransMedTech's model ensures that innovations meet rigorous standards of performance, relevance, usability, cost-efficiency, and societal impact. It also actively mitigates barriers to adoption and commercialization, anchoring innovation in the needs and realities of users and health systems.

This lecture will outline the experience accumulated in the development and operation of the TransMedTech open innovation ecosystem since its inception in 2016. It will examine the foundational principles that informed its implementation, key milestones and achievements, as well as the ecosystem's contributions to

advancing inclusive, sustainable, accessible, and value-based health innovation. The presentation will also highlight the impacts generated across the innovation ecosystem, while offering forward-looking perspectives on how Living Labs such as TransMedTech can continue to play a transformative role in strengthening the medtech innovation landscape through interoperable and socially responsible technologies designed for broad adoption—fostering a more dynamic, connected, and inclusive medtech future.

Tablet Technology Compatible with Functional Magnetic Resonance Imaging: Exploring the Brain Activity Associated with Writing and Drawing Behaviour

Simon J. Graham Ph.D.
Senior Scientist, Sunnybrook Research Institute, Toronto, Ontario, Canada
Professor, Department of Medical Biophysics, Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada
simon.graham@utoronto.ca

Functional magnetic resonance imaging (fMRI) has become a cornerstone of neuroscience research. Developed in the early 1990s, fMRI has substantially increased understanding of how various brain regions and networks engage in mental processing to support human behaviours – in health, as well as in disease and injury. As with all neuroimaging modalities, however, fMRI has methodological limitations that impact how brain activity is recorded and interpreted. One striking consequence is that the vast majority of fMRI studies have focused on the brain activity associated with relatively simple and specific behavioural tasks. Despite this, there is considerable scope remaining to expand fMRI capabilities to encompass complex, naturalistic behaviours, as part of the quest to understand normal and abnormal brain function.

One important case in point is the study of complex motor behaviours associated with writing and drawing – essential forms of human communication and expression. The conventional fMRI experimental approaches, as well as the behavioural stimulus delivery and response recording equipment, have led to a large knowledge gap in the neuroimaging of graphonomic abilities. Despite the sparse fMRI literature on this topic, important developments have been made that encompass medical physics and biomedical engineering, basic neuroscience, and clinical neuroscience with potential translation to medical applications.

This interesting and challenging research space will be reviewed in the current presentation. The narrative encompasses basic aspects of fMRI methodology; historical attempts to conduct writing and drawing studies using fMRI; more modern methods that take advantage of touch-sensitive surfaces used in computerized tablets in combination with augmented reality; as well as examples of tablet-based fMRI applications in diverse areas such as neuropsychological testing, neurosurgery, and neurolinguistics. The narrative will conclude with a look to the future in technological developments of fMRI-compatible tablets, and tablet-based neuroimaging research applications that will shed further light on complex, naturalistic behaviours.

Movement Generation and Drawing in Robotics

Sylvain Calinon
Idiap Research Institute and Ecole Polytechnique Fédérale de Lausanne (EPFL)
sylvain.calinon@idiap.ch

***Abstract*— We present an overview of our research to help robots acquire manipulation and drawing skills by imitation and self-refinement. We advocate frugal learning in our research, where frugality has two goals: 1) learning movement behavior skills from only few demonstrations or exploration trials; and 2) learning only the components of the skill that really need to be learned. Toward this goal, our work investigates the roles of movement primitives, distance fields, optimal control and ergodic control strategies as inductive biases to facilitate movement generation and drawing in robotics.**

Keywords*— *robotics, movement primitives, ergodic control, optimal control, frugal learning

I. Introduction

Despite significant advances in AI, robots still struggle with tasks involving physical interaction. Robots can easily beat humans at board games such as Chess or Go but are incapable of skillfully moving the game pieces by themselves (the part of the task that humans subconsciously succeed in). What makes research in robotics both hard and fascinating is that movement skills are tightly connected to our physical world and to embodied forms of intelligence. In the next sections, an overview of representation and control strategies is presented, with the goal of handling limited training data and uncertainty in robot manipulation and drawing activities.

II. Movement primitives as a superposition of basis functions

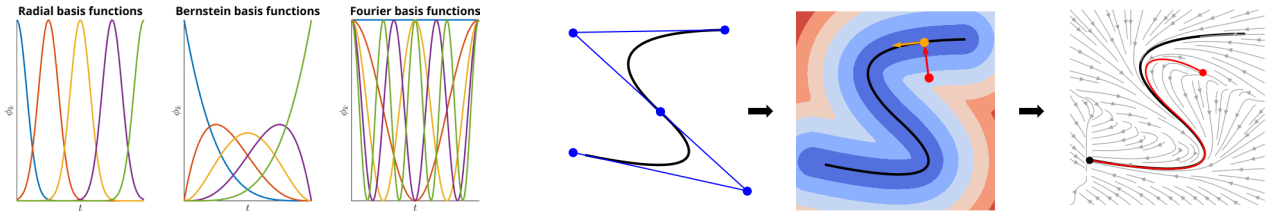


Fig. 1: *Left:* Examples of basis functions to encode movement primitives, see [1] for details. *Right:* Automatic transformation of a piecewise Bézier curve into a distance field and a dynamical system, see [2] for details.

The term *movement primitive* is often used in robotics to refer to the use of basis functions to encode trajectories [1]. It corresponds to an organization of continuous motion signals in the form of a superposition of simpler signals (in parallel and in series), which can be viewed as modular "building blocks" to create more complex movements, see Fig. 1-*left*. This principle, coined in the context of motor control [3], remains valid for a wide range of continuous time signals (for both analysis and synthesis).

The simpler form of *movement primitives* consists of encoding a movement as a weighted superposition of basis functions. This compression/factorization aims at working in a subspace of reduced dimensionality (smaller number of decision variables), while denoising the signal and capturing the essential aspects of a movement (e.g., continuity and smoothness).

Basis functions can for example be used to encode trajectories, whose input is a 1D time variable and whose output can be multidimensional. For basis functions $\phi(t)$ described by a time or phase variable t (see Fig. 1-*left*), the corresponding continuous signal $x(t)$ is encoded with $x(t) = \phi(t)\, w$, with t a continuous time variable and w a vector containing the superposition weights.

When using Bernstein basis functions, instead of considering high-order polynomials to encode long or complicated movement, is is useful in practice to instead consider a concatenation of lower-order polynomials (see Fig. 1-*right* for an example of concatenation with two quadratic Bézier curves). When concatenating curves, constraints on the superposition weights are typically considered. Since these weights can be represented as control points in the case of Bézier curves and splines (see the blue points in Fig. 1-*left*), the last point(s) of a curve and the first point(s) of the next curve will typically be constrained to maintain the continuity of the curve.

This work was supported by the State Secretariat for Education, Research and Innovation in Switzerland and the European Commission's Horizon Europe Program through the INTELLIMAN project (https://intelliman-project.eu/, HORIZON-CL4-Digital-Emerging Grant 101070136) and the SESTOSENSO project (http://sestosenso.eu/, HORIZON-CL4-Digital-Emerging Grant 101070310).

While movement primitives (MPs) provide a compact and modular representation of continuous trajectories, they are essentially time-based or phase-based representations, which can only handle perturbations in a limited way. This motivated the line of research of considering autonomous systems as a time-independent alternative providing control policies based on state-action pairs [4]–[6].

We proposed in [2] a simple and flexible approach that gathers the advantages of both representations by transforming MPs into autonomous systems. The key idea is to transform the explicit representation of a trajectory into an implicit shape encoded as a distance field. This conversion from a time-dependent motion to a spatial representation enables the definition of an autonomous dynamical system with modular reactions to perturbation. Asymptotic stability guarantees are provided by using Bernstein basis functions in the MPs, representing trajectories as concatenated quadratic Bézier curves, which provide analytic distance computation. This approach bridges conventional MPs with distance fields, ensuring smooth and precise motion encoding, while maintaining a continuous spatial representation. By simply leveraging the analytic gradients of the curve and its distance field, a stable dynamical system can be computed to reproduce the demonstrated trajectories while handling perturbations, without requiring a model of the dynamical system to be estimated, see Fig. 1-*right*.

III. Ergodic control

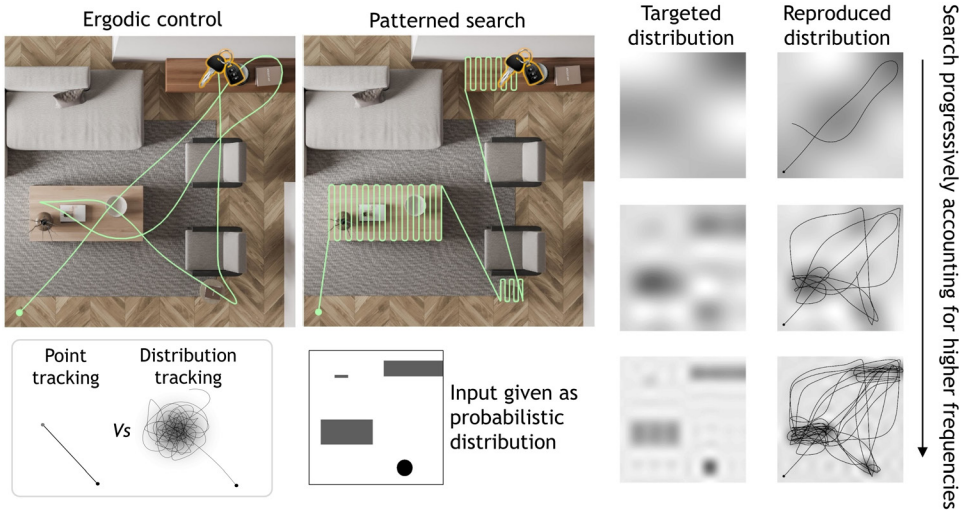


Fig. 2: Ergodic control with a spectral multiscale coverage (SMC) problem formulation.

An accurate reproduction of movements is often not enough to account for all possible sources of uncertainty that can interfere with a robotic task. To handle this issue, ergodic control can be exploited as way to cope with uncertainty of the real world. It can more generally be applied to a wide range of problems requiring the automatic exploration of regions of interest, including active sensing, localization, surveillance, peg-in-hole insertion or construction.

This is particularly helpful when the available sensing information and/or robot actuators are not accurate enough to fulfill the task with a standard controller. This includes situations with limited vision, soft robotic manipulator, or when the relative location of the robot gripper and the grasped tool/object is uncertain (namely, by specifying the task as a distribution instead of a point, see Fig. 2-*bottom*). In all these situations, the robot still has crude information about these locations, which can be exploited for guidance towards most promising areas.

In a collaborative task, ergodic control can also be used when the user's input is not accurate enough to fully reproduce the task, which then requires the robot to explore around the requested input (e.g., a point of interest selected by the operator). For picking and insertion problems, ergodic control can be applied to move in the region of the picking/insertion point, thereby facilitating the prehension/insertion [7]. It can also be employed for active sensing and localization (either detected autonomously, or with help by the operator). Here, the robot can plan movements based on the current information density, and can recompute the commands when new measurements are available (i.e., updating the spatial distribution used as target).

Research in ergodic control addresses fundamental challenges linking machine learning, optimal control, signal processing and information theory. Mathematically, ergodic theory studies the connection between the time-averaged and space-averaged behaviors of a dynamical system, which enables optimal exploration of an information distribution. A conventional tracking problem in robotics is characterized by a target to reach, requiring a controller to be computed to reach this target. In *ergodic control*, instead of providing a single target point, a probability distribution is given to the robot, which must cover the distribution in an efficient way. Ergodic control thus consists of moving within a spatial distribution by spending time in each part of the distribution in proportion to its density. The term ergodicity corresponds here to the difference between the

time-averaged spatial statistics of the agent's trajectory and the target distribution to search in. By minimizing this difference, the resulting controller generates natural exploration behaviors.

Several ergodic control problem formulations exist, including *spectral multiscale coverage* (SMC) [8], diffusion-based [9], energy-based [10] and kernel-based [11] approaches. It can also be applied to surfaces characterized by meshes or point clouds [12]. The objective of SMC corresponds to a tracking problem in the spectral domain (matching of frequency components), see Fig. 2-*bottom*. The advantage of this simple control formulation is that it can be easily combined with other control objectives and constraints. It requires the spatial distribution to be decomposed as Fourier series, with a cost function comparing the spectral decomposition of the robot path with the spectral decomposition of the distribution, in the form of a weighted distance function with a decreasing weight from low frequency to high frequency components. The resulting controller allows the robot to explore the given spatial distribution in a natural manner, by starting from a crude exploration and by refining the search progressively (i.e., matching the Fourier coefficients with an increasing importance from low to high frequency components), see Fig. 2-*right*.

Fig. 2-*left* illustrates the use of ergodic control to search for a key in a living room, where regions of interest are specified in the form of a distribution over the spatial domain (see *bottom-center* inset). These regions of interest correspond to the different tables in the room where it is likely to find the key. The left and right columns respectively depict an ergodic search and a patterned search. For a coverage problem in which the duration is given in advance, a patterned search is well suited because the agent will optimize the path to reduce the number of transitions from one area to the other. In contrast, an ergodic control search will typically result in multiple transitions between the different regions of interest. When searching for a key, since we are interested in finding the key as early as possible, we have a problem in which the total search duration cannot be specified in advance. In this problem setting, an ergodic search is better suited, generating a natural way of exploring the environment by first proceeding with a crude search in the whole space and by progressively fine-tuning the search with additional details until the key is finally found. Indeed, to search for a key, a patterned search would be an inefficient and unnatural way of scanning the whole environment regularly by looking at all details regions by regions, instead of treating the problem at different scales (i.e., at different spatial frequencies as in SMC). Insertion tasks can similarly be achieved robustly with ergodic control, yielding a control strategy capable of coping with limited or inaccurate sensing information, see [7] for details.

IV. Application: Drawing robots

The modular movement representations and the control strategies presented in the previous sections are also exploited in our work to provide robots with drawing and writing capabilities.

In Fig. 3-*(a)*, movement primitives were used to generate the upper-body movement of a mini-humanoid robot drawing portraits (incl. bimanual gestures with natural-looking oscillations) [13]. Different shades of gray were generated through concentric arc circles to emulate the movement of a wrist during hatching.

In Fig. 3-*(b)*, a torque-controlled full-size humanoid was used to generate rapid and compliant movements to produce calligraphic graffiti art [14], [15]. In this experiment in collaboration with researchers from Goldsmiths University of London, an interface for artists was developed to let them edit trajectory paths by also considering variations in the underlying dynamics of the traces. To do this, the notion of control points in splines was extended to the use of Gaussian distributions to provide coordination and variation information. Optimal control was then used as the underlying mechanism to generate dynamic motions.

Fig. 3-*(c)* exploits a diffusion-based ergodic control method called HEDAC to draw portraits with a series of small strokes [16]. The underlying optimal control problem consists of covering a given distribution (the grayscale portrait image), so that the robot will pass more frequently on the dark regions of the image, effectively producing different shades of gray with a single ballpoint pen. When observing the robot at work, the doodles appear meaningless at first sight, with the global objective progressively revealed as time passes. Interestingly, while the result might appear to be composed of aleatory strokes, the whole process is purely deterministic and does not introduce any stochasticity in the generative process (namely, the code does not involve any random number generation). These doodles appear to be random when they are observed from close distance, but they are in reality the result of the optimal control problem, with the doodles revealing the global objective only when the drawing is finished.

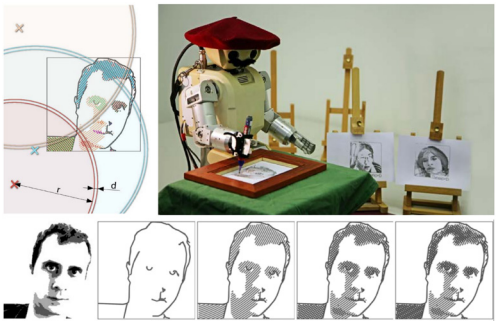
References

[1] S. Calinon, "Mixture models for the analysis, edition, and synthesis of continuous time series," in *Mixture Models and Applications* (N. Bouguila and W. Fan, eds.), pp. 39–57, Springer, 2019.

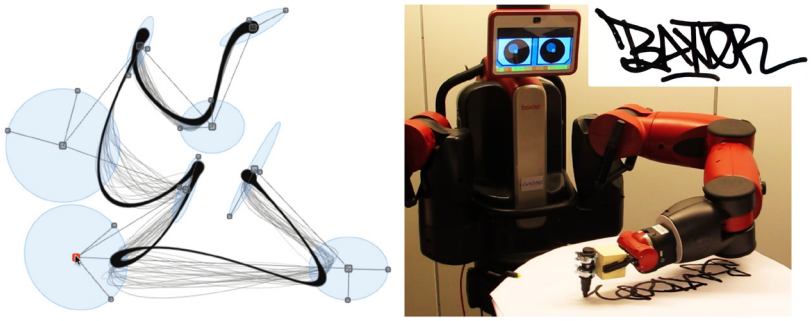
[2] Y. Li and S. Calinon, "From movement primitives to distance fields to dynamical systems," *arXiv:2504.09705*, 2025.

[3] F. A. Mussa-Ivaldi, S. F. Giszter, and E. Bizzi, "Linear combinations of primitives in vertebrate motor control," *Proc. National Academy of Sciences*, vol. 91, pp. 7534–7538, August 1994.

[4] M. Hersch, F. Guenter, S. Calinon, and A. G. Billard, "Dynamical system modulation for robot learning via kinesthetic demonstrations," *IEEE Trans. on Robotics*, vol. 24, no. 6, pp. 1463–1467, 2008.



(a) Salvador DaBot [13]



(b) Dynamic tags writing with Baxter [14], [15]



(c) DrozBot [16]

Fig. 3: Experiments with various forms of drawing robots.

[5] S. M. Khansari-Zadeh and A. Billard, "Learning control Lyapunov function to ensure stability of dynamical system-based robot reaching motions," *Robotics and Autonomous Systems*, vol. 62, no. 6, pp. 752–765, 2014.
[6] M. Saveriano, "An energy-based approach to ensure the stability of learned dynamical systems," in *Proc. IEEE Intl Conf. on Robotics and Automation (ICRA)*, pp. 4407–4413, 2020.
[7] S. Shetty, J. Silvério, and S. Calinon, "Ergodic exploration using tensor train: Applications in insertion tasks," *IEEE Trans. on Robotics*, vol. 38, no. 2, pp. 906–921, 2022.
[8] G. Mathew and I. Mezić, "Spectral multiscale coverage: A uniform coverage algorithm for mobile sensor networks," in *Proc. IEEE Conf. on Decision and Control*, pp. 7872–7877, Dec 2009.
[9] S. Ivić, B. Crnković, and I. Mezić, "Ergodicity-based cooperative multiagent area coverage via a potential field," *IEEE Trans. on Cybernetics*, vol. 47, no. 8, pp. 1983–1993, 2017.
[10] A. Seewald, C. J. Lerch, M. Chancán, A. M. Dollar, and I. Abraham, "Energy-aware ergodic search: Continuous exploration for multi-agent systems with battery constraints," in *Proc. IEEE Intl Conf. on Robotics and Automation (ICRA)*, pp. 7048–7054, 2024.
[11] M. M. Sun, A. Gaggar, P. Trautman, and T. Murphey, "Fast ergodic search with kernel functions," *IEEE Transactions on Robotics*, vol. 41, pp. 1841–1860, 2025.
[12] C. Bilaloglu, T. Löw, and S. Calinon, "Tactile ergodic coverage on curved surfaces," *IEEE Transactions on Robotics (T-RO)*, vol. 41, pp. 1421–1435, 2025.
[13] S. Calinon, J. Epiney, and A. G. Billard, "A humanoid robot drawing human portraits," in *Proc. IEEE-RAS Intl Conf. on Humanoid Robots (Humanoids)*, pp. 161–166, 2005.
[14] D. Berio, S. Calinon, and F. Fol Leymarie, "Learning dynamic graffiti strokes with a compliant robot," in *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems (IROS)*, pp. 3981–3986, 2016.
[15] D. Berio, F. Fol Leymarie, and S. Calinon, "Interactive generation of calligraphic trajectories from Gaussian mixtures," in *Mixture Models and Applications* (N. Bouguila and W. Fan, eds.), pp. 23–38, Springer, 2019.
[16] T. Löw, J. Maceiras, and S. Calinon, "drozbot: Using ergodic control to draw portraits," *IEEE Robotics and Automation Letters (RA-L)*, vol. 7, no. 4, pp. 11728–34, 2022.



Session

Biomarkers

Lognormality-Based Aggregation to Assist Kindergarten Pupil's Graphomotor Skills Analysis

C REMI [1], S REGIS [1], J VAILLANT [1], R PLAMONDON [2]

[1] *LAMIA Pôle Guadeloupe, Université des Antilles, 97110, Pointe-à-Pitre Cedex,* [2] *Laboratoire Scribens, Génie Électrique, Polytechnique Montréal, 2500 Chemin de Polytechnique, Montréal, H3T-1J4, Québec, Canada*

celine.remi@univ-antilles.fr, sebastien.regis@univ-antilles.fr, jean.vaillant@univ-antilles.fr, rejean.plamondon@polymtl.ca

***Abstract*— We introduce lognormality-based aggregated features as potential biomarkers to help teachers' early identification of pupils needing clinical evaluations. This ongoing study indicates that aggregated SNR (Signal-to-Noise Ratio) can track pupils' kinematic performances progression, regression or stagnation. To link the objective SNR metrics to the subjective teacher evaluation we use a fuzzy logic approach. Correlation tests show that this feature complement teachers' assessments and could help in neuromotor control evaluation.**

Keywords—aggregation, lognormality principle, kindergarten, graphomotor skills, pseudo-letter, biomarker, fuzzy sets, fuzzy logic

I. INTRODUCTION

Teaching handwriting is complex and requires well-trained teachers. In the French education system, teachers play a crucial role in identifying students with behavioural and psychomotor difficulties, as detailed in [1]. These issues may include neurodevelopmental disorders affecting motor coordination and fine motor skills. Early detection of these problems is challenging, often taking years before being properly understood by families or medical professionals. The average age for diagnosing developmental coordination disorder (DCD) is 7.3 years, which is too late to prevent further complications, as reported by the French public senate [2]. Teachers, especially in kindergarten, are essential in recognizing children who need early medical or multidisciplinary intervention. Delayed identification can lead to significant socio-emotional harm to the child, their family, and the teacher.

Various digital tools such as Kaligo [3], Copilotr@ce [4], and others [5] can support teachers in training graphomotricity, handwriting, or numerous other skills based on graphomotor gestures. While these tools capture graphomotor gestures and extract objective data on visuo-spatial, kinaesthetic, procedural, and kinematic levels—data that teachers may find difficult to manage in large classes—they are not designed to assist teachers in early warning detection. Additionally, these tools do not consider the graphomotor aspects of graphomotor gestures in their automated evaluations. Scientific evidence indicates that kinematic data can be useful in identifying behavioural differences related to graphomotor issues in movement planning and execution, which could serve as warning signals or progression biomarkers [6]. In [5], an alert assistance system based on automated sigma-lognormal analysis of pupils' graphomotor productions was described. This system collects graphomotor traces over time to estimate warning indices using lognormality principles [7]. It transparently analyzes students' work within educational activities assigned by teachers, differing from clinical devices like the Lognometer [8], which use specific psychometric tests. As shown by 5 sigma-lognormal modelling in figure 1, previous research with activities such as drawing scribbles, copying letters, and writing words supports the effectiveness of the Kinematic Theory and its sigma-lognormal parameters in reconstructing young pupils' handwriting movements.

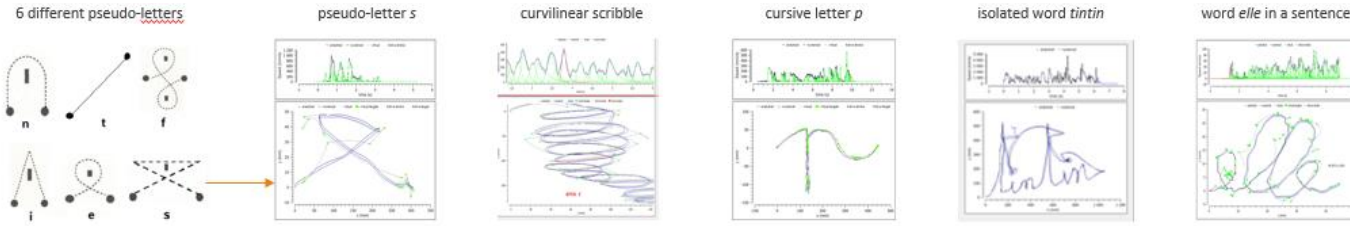


Fig. 1. 6 pseudo-letters collected in project "Is there a co-pilot on board?" the beginning points are the ones left and 5 examples of sigma-lognormal reconstructions.

This paper is part of an ongoing work to implement and validate a lognormality-based alert tool for teachers inspired by [5]. It aims at determining if the aggregated values of global sigma-lognormal parameters computed from pupils' pseudo-letters workflows usually produced during the teaching process could provide useful knowledge to teachers' practices. We have chosen to work with pseudo-letters because their handwriting trajectories, that are a mix of curvilinear or rectilinear strokes between a beginning and an end point (fig. 1), are like those used in [8]. First, we describe the raw data used, introduce the aggregation approach and test it.

Next, we discuss the use of fuzzy subsets theory and aggregation operators to study correlation between teachers' assessments and the aggregated sigma-lognormal parameters. The relevance of our aggregation approach is analysed before the conclusion.

II. Raw graphomotor workflows investigated in this ongoing work

Our present raw graphomotor workflows consisted in 67 subsets of a total of 3127 trajectories collected during two campaigns organized in March and June. All these trajectories are extracted from the original dataset generated by the project "Is there a co-pilot on board?" [9,10]. Three kindergarten grades of pupils took part in this project: PS (Petite Section, first year of kindergarten), MS (Moyenne Section, second year of kindergarten), GS (Grande Section, third year of kindergarten). During this project, pupils' teachers and families conducted sessions of graphomotor pseudo-letter training with fingers and markers over four months from March to June. Teachers provided anonymized qualitative evaluations for each pupil who took part in this project: once in March before the project started and once in June after the training period. The trajectories were created with a Wacom variable pressure pen with a non-inking tip by voluntary pupils on a Wacom tablet. Next, they were reconstructed using sigma-lognormal modelling. The raw graphomotor workflows, are composed with only trajectories with SNR (Signal-to-Noise Ratio) values greater than 15 dB. For this study, we consider 5 groups of raw graphomotor workflows. Each of these groups are defined by a couple (Grade, month). For each of these 5 groups, grade corresponds to the scholar level for pupils whose produced workflows, while month defines the campaign. Table 1 provides the numbers of raw graphomotor workflows, the pupils' sex, their average age, and the standard deviation age for each group. The reader will note that these 67 workflows were produced by 51 pupils.

TABLE I. Numbers of pseudo-letters workflows, with their producers' average ages, standard deviation ages, grades, periods and sex for each of the 5 groups.

Grade-Month	PS-March	MS-March	MS-June	GS-March	GS-June	GLOBAL
Number of raw graphomotor workflows	5	13	10	29	10	67
Pupils' sex (F/M)	3/2	8/5	7/3	10/19	1/9	
Pupils' Average age (in month)	43,82	56,98	65,45	67,72	71,24	
Pupils' Standard deviation of age	3,58	3,92	5.75	4,98	3,45	

III. Approach for the aggregation of the raw sigma-lognormal parameters

A. An aggregation approach based on the lognormality principle

According to the principle of convergence towards lognormality outlined in [9], each task that a student encounters—whether in class with their teacher, at home with their family, or as part of the experimental collection of their traces on the digitizer—represents a challenging situation along their training process. Indeed, they must efficiently mobilize their graphomotor and visuo-spatial faculties to perform each proposed task. Viewed in this manner, these multiple trainings can be seen as opportunities for a pupil to engage, progress, update, and demonstrate the level of his graphomotor abilities. Based on previous works, it is known that: (a) the Kinematic Theory is not overly sensitive to the noise induced by the conditions of acquisition of trajectories within kindergarten [10], and consequently, (b) the trend towards lognormality of fine movement control is accompanied by observations that SNR values progressively increase according to grade because movements that become better controlled are better modelled [10].
We hypothesize that higher levels and degrees of consolidation of graphomotor control should make this phenomenon of migration towards normality more perceptible for students, according to their age in months and for each grade according to the period of observation. This hypothesis considers the aggregation of all the SNR values estimated for each production of pseudo-letters workflow. Ideally, this increase should reach a level compatible with the maximum graphomotor control achievable for each age or grade level. We have chosen the couple of usual aggregation functions cumulated mean SNR and cumulated median SNR to aggregate the estimated SNR values during a periodically recorded workflow for a student or a group of students tracked by the instructor. The choice of these two simple functions for measuring central tendency [11] is based on the following considerations: They are sensitive and exhibit a trend towards lognormality based on training duration, even across various pseudo-letters in the absence of developmental issues. The difference between mean and median values and their change over time serve as indicators of the level and ability to consolidate graphomotor control in the subjects.

B. Analysis of the behaviour of the aggregated SNR

Based on the principle of migration towards lognormality, to verify the relevance of the Aggregated SNR, we have estimated the means and medians values for each of the 51 pupils' workflows recorded online and by grade for each March and June campaign.

Figures 2a and 2b show the behaviours of the features SNR_mean and SNR_median respectively according to the couple (grade, period) and to pupil ages. The boxplots in figure 2a are ordered according to the average ages provided in table 1.

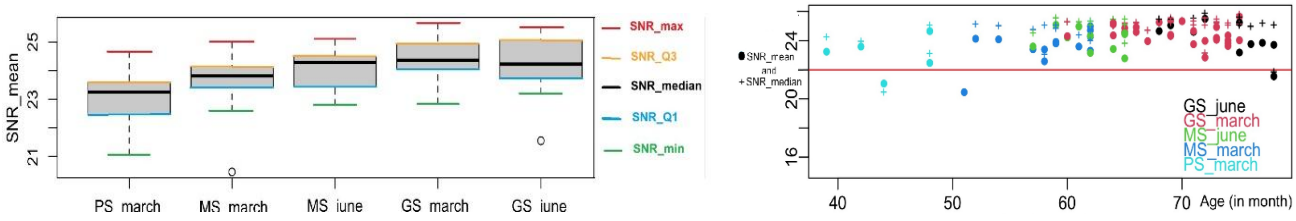


Fig. 2. (a) at left shows min (except atypical individuals), max, median, first and third quartiles of SNR_mean while (b) at right shows SNR_mean (•) and SNR_median (+) values for each pupil. The red line is an empirical threshold to separate atypical pupils from others with respect to the teachers' opinions.

Figure 2a indicates that higher education levels and longer exposure to graphic design correlate with increased SNR_mean (Kruskal-Wallis test: p-value = 0.013 for couple (grade, period)). It also shows a median ceiling effect over the first three years of schooling. This suggests that, according to the lognormality principle, the automated estimation of at least SNR_mean and SNR_median when performing pseudo-letters might provide useful information on kindergarten pupils' level of graphomotor control useful to complete assessments and to monitor the evolution of the apprenticeship. The correlation between SNR_mean and age is 0.33 (*p. value=0.006*), while between SNR_median and age it is 0.37 (*p. value=0.002*), both significantly positive.

IV. Comparison of Aggregated SNR with Teachers' Usual Academic Assessments

Kindergarten teacher assessments can vary for a single exercise, often using letters A (very satisfactory), B (satisfactory), C (fair), D (insufficient), and E (very insufficient). In the project "Is there a co-pilot on board?", students were evaluated on graphism, handwriting, and engagement. Three students—one from PS, MS, and GS—stood out due to their low SNR_mean and SNR_median (see fig. 2b, the three dot under the red threshold line, two of them are statistical outliers while the third one has the smaller SNR values in its group PS-march). Indeed, the PS students' evaluations were fair or insufficient. The MS and GS students received B grades but had additional comments indicating their performance was lower than suggested by letter B (see Table 2).

TABLE II. Teachers' Scoring and Comments for the 3 Pupils who had the Lowest SNR_mean.

Pupil's grade	PS	MS	GS
Graphism/ Handwriting/ Interest scoring	C/D/C	B/B/B	B/B/B
Comment		begins to master the sense of writing	was not very neat and rather expeditious in writing and graphic design but is starting to make efforts

In this context, the next step was to determine if the knowledge provided by the global statistical feature Aggregated SNR was similar or not to those provided by such usual type of teachers' qualitative evaluation of pupils' graphomotor performances at school. To compare quantitatively the results of teachers' evaluations with the aggregations of the SNR and the SNR/NbLog ratios we have considered the qualitative and multiple nature of assessments using fuzzy set theory and its aggregation operators.

A. Fuzzy subsets and aggregation operators

Fuzzy set theory was developed to address the inherent imprecision in human cognition and knowledge representation. Consequently, fuzzy subsets provide a suitable framework for modelling subjective and qualitative assessments made by human evaluators. It is important to remember that a classical subset A of X is characterized by a function χ that assigns the value 0 to elements of X that do not belong to A, and 1 to elements that do belong to A [12]:

$$\chi_A: X \rightarrow \{0,1\} \quad (1)$$

In contrast, a fuzzy subset B of X is defined by a membership function f that assigns each element x of X a degree $f(x)$ between 0 and 1, representing the extent to which x belongs to B [10]:

$$f_B: X \rightarrow [0,1] \quad (2)$$

Within the framework of fuzzy logic, aggregation operators have been designed to synthesize fuzzy information derived from multiple sources. These operators exhibit a wide range of properties, enabling the modelling of diverse real-world scenarios. Notably, these operators, initially developed for fuzzy sets, are applicable to any domain where the underlying data can be normalized within the unit interval [0, 1] (see for instance [13]). Aggregation operators can be categorized based on their behaviour with respect to the minimum and the maximum of the input set. Specifically, we can distinguish four classes:

- Conjunctive operators, exemplified by t-norms and copulas, are characterized by the property that their output is always less than or equal to the minimum value within the aggregated set.
- Disjunctive operators, like t-conorms, conversely, produce results always greater than or equal to the maximum value.
- Averaging operators, including the arithmetic mean or Ordered Weighted Averaging (OWA) operators, yield results that lie strictly between the minimum and maximum values of the input set. This behaviour aligns with their function of representing a compromise or central tendency within the data.

- Hybrid operators encompass all aggregation functions that do not fall into the categories above. This class includes, but is not limited to, mixed connectives, uninorms, nullnorms, and symmetric sums, thereby representing a broader range of aggregation behaviours beyond the basic conjunctive, disjunctive, and averaging types.

The choice of the type of aggregation operator depends on the application and the objectives of the data fusion.

B. Application for fuzzification and aggregation of teacher evaluations

We have fuzzified the teachers' evaluations by using trapezoidal membership functions, as shown in Figure 3 [14,15].

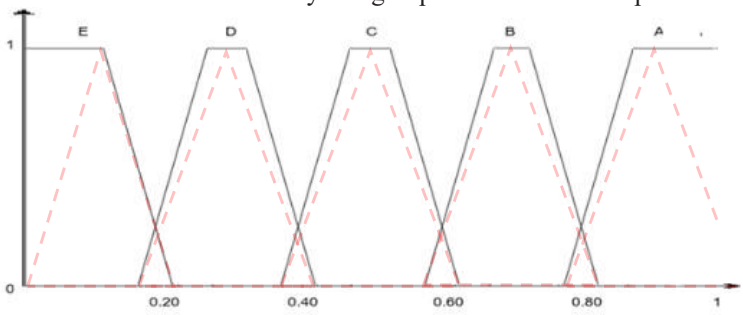


Fig. 3. Fuzzy membership functions of teachers' qualitative assessments (A: very satisfactory; B: satisfactory; C: fair; D: insufficient; E: very insufficient).

Next, to merge the fuzzified evaluations, we have chosen 3 different average type operators, which were the most suitable type of aggregation operators to make a compromise between evaluation values carried out by the same teacher. They are:

- the classical arithmetic average (AA):

$$\forall\, x_i \in [0,1],\ i = 1,\cdots,n$$

$$AA(x_1,\cdots,x_n) = \frac{1}{n}\Sigma_n^{i=1}\, x_i \quad (3)$$

- the triple average Pi (A3P) which is more discriminating than the arithmetic average [16]:

$$\forall x_i \in [0,1],\ i = 1,\cdots,n$$

$$A3P(x_1,\cdots,x_n) = \frac{1}{1+\Pi_{i=1}^{n}\left(\frac{1-x_i}{x_i}\right)^{1/n}} \quad (4)$$

- one of the OWA averages proposed by Filev and Yager [17]. This OWA has a parameter that can be varied in [0,1]:

$$\forall x_i \in [0,1],\ i = 1,\cdots,n$$

$$OWA(x_1,\cdots,x_n) = \Sigma_{i=1}^{n} w_i\, y_i \quad (5)$$

where y_i is the ith largest element of the set of the x_i and:

$$w_i = \alpha \text{ with } \alpha \in [0,1] \ \ if\ i = 1\ ;\ w_i = w_{i-1}.(1-w_1)\ \ if\ i = 2,\cdots,n-1\ \ ;\ w_i = w_{i-1}.(1-w_1)/w_1\ if\ i = n$$

C. Results of the correlation tests aggregation of teachers' opinions on their pupils' skills

Figure 4 shows that the A3P aggregatioqn of fuzzified handwriting and interest teachers' opinions (A3P) does not have a correlation with SNR_mean (Correlation Coefficient 0.139, p-value=0.262). There is no significant correlation between SNR/NbLog_median and A3P (Correlation Coefficient 0.232, p-value=0.059) as indicated in Figure 4. However, there is a low significant correlation between A3P and both SNR_median (Correlation Coefficient 0.246, p-value=0.045) and SNR/NbLog_mean (Correlation Coefficient 0.247, p-value=0.044). These low correlations do not suggest that the information provided by the sigma-lognormal approach is redundant with the teacher's qualitative evaluations. On the contrary, the information given by the sigma-lognormal model, as shown in the case of the 3 students reported in Table II, appears to additionally support the findings made by the teacher regarding the mastery of graphomotor gestures. However, these observations did not lead the teacher to unfavourably evaluate the overall performance of their students for the evaluation period concerned.

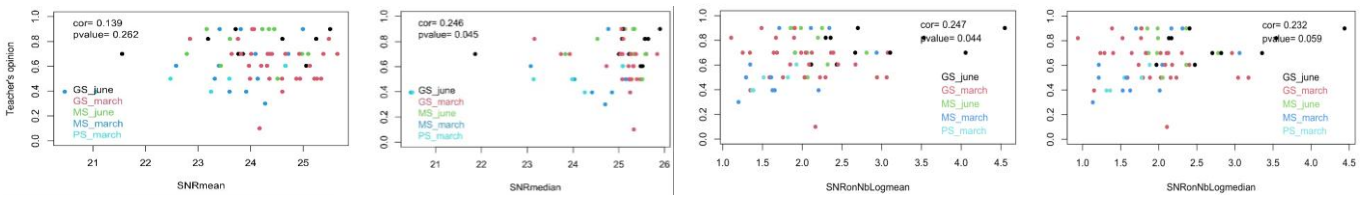


Fig. 4. From left to right, correlation of teacher's opinions with: SNR-mean; SNR-median, SNR/NbLog mean SNR/NbLog-median.

V. Conclusion

We propose lognormality-based aggregated features as biomarkers to help teachers identify students needing clinical evaluation. Our study shows that Aggregated SNR can indicate changes in pupils' kinematic performances based on their age and grade. Correlation tests reveal they complement teachers' assessments without redundancy, leading to more accurate evaluations. These first results of this ongoing work may need further investigations. However, observations made for Aggregated SNR suggest that a lognormality-based approach may be relevant regarding the teachers' quantifying of young Pupils' control of graphomotor movements and their temporal evolutions between 3 and 6 years old during pseudo-letter training through iterative online acquisitions and sigma-lognormal analysis. Teachers could use it to compare individual pupil performance with group trends and to identify those in long-term or non-permanent difficulty. This estimation could be done independently of the pseudo-letter types.

References

[1] HAS (2020). Argumentaire « Troubles du neuro-développement - Repérage et orientation des enfants à risque », HAS, 226 pp.

[2] Romagny, A.-S. (2024), RAPPORT FAIT au nom de la commission des affaires sociales sur la proposition de loi visant à améliorer le dépistage des troubles du neurodéveloppement, l'accompagnement des personnes qui en sont atteintes et le répit de leurs proches aidants, Sénat, n°245, 83 pp.

[3] Harbaum T. et al., (2024). KIHT: Kaligo-Based Intelligent Handwriting Teacher; 2024 Design, Automation & Test in Europe Conference & Exhibition (DATE), Valencia, Spain, pp. 1-6.

[4] Rémi, C., Nagau, J. (2022). Copilotrace: a platform to process graphomotor tasks for education and graphonomics research. In: Carmona-Duarte, C., Díaz, M., Ferrer, M.A., Morales, A. (eds.) Intertwining Graphonomics with Human Movements - 20th International Conference of the International Graphonomics Society, IGS 2021, Las Palmas de Gran Canaria, Spain, 7–9 June 2022, Proceedings. LNCS, vol. 13424, pp. 129–143.

[5] Plamondon, R., Pirlo, G., Anquetil, É., Rémi, C., Teulings, H.-L., & Nakagawa, M. (2018). Personal digital bodyguards for e-security, e-learning and e-health: A prospective survey. Pattern Recognition, 81, 633-659.

[6] Plamondon, R., et al (2023) Lognormality: An Open Window on Neuromotor Control. in Parziale, A., Diaz,M., Melo, F.Graphonomics inHuman Body Movement: Bridging Research and Practice from Motor Control to Handwriting Analysis and Recognition. Proceedings 21st International Conference of the International Graphonomics Society, IGS2023.Spronger Lecture Notes in Computer Sciences p(14285): 205-258, Springer Nature Switzerland

[7] Plamondon, R. (2021). The Lognormality Principle: A Personalized Survey,. In Plamonddon, R., Marcelli, A., Ferrer-ballester, M.A., (Eds) The Lognormality Principle and its Applications in e-Security, e-Learning and e-Health World Scientific Publishing Series in Machine Perception and Artificial Intelligence (Vol. 88, p. 1-39).

[8] O'Reilly, C., Plamondon, R., & Faci, N. (2022). The Lognometer A New Normalized and Computerized Device for Assessing the Neurodevelopment of Fine Motor Control in Children, 26th International Conference on Pattern Recognition (ICPR 2022), Montreal, CANADA,. paper 952-058.

[9] Plamondon, R., O'Reilly, C., Rémi, C., & Duval, T. (2013). The lognormal handwriter: learning, performing, and declining. Frontiers *in Psychology*, *4*.

[10] Duval, T., Rémi, C., Plamondon, R., Valliant, J., & O'Reilly, C. (2015). Combining sigma-lognormal modelling and classical features for analysing graphomotor performances in kindergarten children. Human Movement Science, 43, 183-200.

[11] Grabisch M., Marichal J.-L., Mesiar R., Pap E., Aggregation functions: Means. Information Sciences, 2011, 181 (1), pp.1-22.

[12] Zadeh, L. A. (1965). Fuzzy sets. Information and Control.

[13] Plocoste, T., Regis, S., Nuiro, S. P., & Sankaran, A. (2024). Application of aggregation operators for forecasting PM10 fluctuations: From available Caribbean data sites to unequipped ones. Atmospheric Pollution Research, 15(6), 102116.

[14] Hameed, I. A., & Sorensen, C. G. (2010). Fuzzy systems in education: a more reliable system for student evaluation (pp. 978-953). ISBN.

[15] Khan, A. R., Amin, H. U., & Rehman, Z. U. (2011). Application of expert system with fuzzy logic in teachers' performance evaluation. International Journal of Advanced Computer Science and Applications, 2(2).

[16] Emilion, R., Regis, S., & Doncescu, A. (2013). A general version of the triple π operator. International journal of intelligent systems, 28(10), 957-973.

[17] Filev DP, Yager RR. On the issue of obtaining OWA operator weights. Fuzzy Set Syst 1998;94:157–169.

Handwriting Analysis for the Detection of Lewy Body Dementia: Methods, Challenges, and Perspectives

Tiziana D'Alessandro, Claudio De Stefano, Francesco Fontanella, Emanuele Nardone,
Olga Pustovalova, Alessandra Scotto di Freca
Università degli Studi di Cassino e del Lazio Meridionale, Via Gaetano di Biasio, 43, 03043 Cassino, FR, Italy
{tiziana.dalessandro, destefano, fontanella, emanuele.nardone, olga.pustovalova, a.scotto}@unicas.it

***Abstract*— Lewy Body Disease (LBD) is a progressive neurodegenerative disorder comprehending Dementia with Lewy Bodies (DLB) and Parkinson's Disease Dementia (PDD). These conditions are characterized by cognitive impairment, motor dysfunction, and the accumulation of abnormal aggregates of alpha-synuclein proteins (Lewy bodies) in the brain. DLB is diagnosed when cognitive symptoms appear before or within one year of motor symptoms. In contrast, PDD is identified w hen d ementia d evelops a t l east o ne y ear a fter t he o nset o f Parkinsonian motor symptoms. An early and accurate diagnosis of LBD remains a major clinical challenge, as its symptoms overlap with other dementias, such as Alzheimer's Disease (AD). Handwriting analysis is a promising non-invasive tool for detecting LBD-related motor and cognitive impairments. Digital tablets enable the precise acquisition of handwriting kinematics, allowing for the quantification o f fi n e mo to r control deficits. S tudies s uggest t hat L BD p atients e xhibit d istinct h andwriting a lterations, s uch a s m icrographia, i ncreased p auses, r e duced fluidity, and irregular pressure distribution. Advanced artificial i ntelligence (A I) t e chniques, p a rticularly m a chine l e arning (M L) a n d d e ep learning (DL), have been employed to extract meaningful patterns from handwriting data. Supervised and unsupervised learning methods have shown promising results in differentiating LBD from other neurodegenerative disorders. However, challenges remain in terms of data heterogeneity, model explainability, and the need for vast and longitudinal datasets. Future research should focus on integrating multimodal data, improving algorithm robustness, and ensuring clinical interpretability. This paper explores the current state of handwriting-based LBD diagnostics, highlighting the potential of AI-driven approaches in improving early detection and disease monitoring.

Keywords*— *Neurodegenerative Disorder, Lewy Body Disease, Handwriting Analysis, Artificial Intelligence

I. Introduction

Lewy Body Disease (LBD) is a neurodegenerative disorder characterized by the abnormal accumulation of α-synuclein aggregates in brain cells, called Lewy bodies. This disease can be categorized into two subtypes: Dementia with Lewy Bodies (DLB) and Parkinson's Disease Dementia (PDD). These conditions share the same neuropathological characteristics but differ in the timing of the symptoms onset: DLB is diagnosed when cognitive impairment appears before or within one year of motor symptoms; instead, PDD is identified when dementia develops at least one year after the onset of Parkinsonian motor symptoms.

DLB is characterized by progressive cognitive decline, particularly affecting visuospatial skills, attention, executive function, and memory, usually with at least one symptom of parkinsonism, such as bradykinesia, gait impairment, or tremor. Other symptoms include fluctuations in attention and alertness, detailed visual hallucinations (affecting $\approx 80\%$ of patients), and REM sleep behaviour disorder (76% of cases). Diagnosing DLB correctly is critical, as it shares features with Alzheimer's Disease (AD), and misdiagnosis can lead to severe adverse reactions to antipsychotic medications, which are commonly prescribed for AD. In terms of occurrence statistics, DLB is the second most common type of dementia, covering 10%–15% of dementia cases, although timely diagnosis is still an open issue. Similarly, LBD and Parkinson's Disease (PD) are closely related: most individuals with PD exhibit Lewy bodies in their brain, and 25%–30% of PD patients develop dementia, with prevalence rising to 83% after 20 years of disease progression. Given the overlap between DLB and PDD (which share 85% of clinical symptoms), early and accurate diagnosis is essential, though particularly challenging. Identifying prodromal signs—such as REM sleep behaviour disorder—could help estimate the risk of LBD progression before a significant cognitive decline occurs.

Thus, diagnosing LBD is challenging because of the overlap of symptoms with other neurodegenerative diseases. This complexity highlights the need for reliable biomarkers to enable early and differential diagnosis. In this context, handwriting analysis is considered a promising and non-invasive diagnostic tool capable of detecting motor and cognitive impairments associated with LBD. Advanced technologies, such as digital tablets, allow for the collection of detailed kinematic data (e.g., speed, pressure, fluency, pauses, and in-air movements), providing quantitative insights that may help distinguish DLB from PDD and other forms of dementia. This paper explores the role of handwriting analysis in the diagnosis of LBD, focusing on key assessment parameters, analytical methodologies, and the potential of artificial intelligence in processing handwriting-based biomarkers.

II. LBD Diagnosis

The diagnosis of LBD, including both its categories, DLB and PDD, is a complex process that requires a multimodal approach taking into account clinical evaluation, cognitive testing, neuroimaging, and biomarker analysis. Because some

symptoms of LBD overlap with AD or PD, an accurate and timely diagnosis is crucial to start appropriate treatment and avoid adverse reactions to medications, particularly antipsychotics, which can be harmful to patients with LBD. In conclusion, LBD diagnosis demands a combination of clinical expertise, standardized testing, neuroimaging, and emerging digital tools. Integrating handwriting analysis and other digital biomarkers into the diagnostic process can improve early detection and differentiation of related neurodegenerative diseases, allowing for better patient care and treatment strategies.

A. Clinical Evaluation

The diagnostics of LBD begins with a clinical assessment, including an analysis of medical history followed by an evaluation of cognitive, motor, and neuropsychiatric symptoms. On the one hand, DLB is distinguished by early cognitive impairment, particularly in visuospatial abilities, attention, and executive function, and hallmark symptoms such as fluctuating alertness, recurrent visual hallucinations, REM sleep behaviour disorder (RBD), and Parkinsonian motor signs. On the other hand, the onset of PD motor symptoms in PDD is at least one year before dementia, with dementia developing progressively over time.

B. Cognitive Testing

Clinicians require standardized neuropsychological tests to assess cognitive decline and distinguish LBD from other dementias [1], [2]. The cognitive assessment usually comprises questionnaires and the execution of tasks, such as drawing, writing, or speaking. Commonly employed tools include the Mini-Mental State Examination (MMSE), which assesses cognitive impairment. The Clock Drawing Test (CDT) evaluates the visuospatial and executive functions, which are impaired in DLB. The Trail Making Test (TMT) measures cognitive flexibility and processing speed. The Neuropsychiatric Inventory (NPI) evaluates behavioural and psychiatric symptoms, such as hallucinations, depression, and agitation. The Unified Parkinson's Disease Rating Scale (UPDRS) measures motor and non-motor symptoms, including cognitive and functional impairments associated with PD. Additionally, the Dementia Cognitive Fluctuations Scale (DCFS) is used to assess variations in attention and alertness, a common characteristic of DLB.

C. Neuroimaging

Advanced brain imaging techniques can help detect neuropathological changes associated with LBD. Magnetic Resonance Imaging (MRI) is useful for assessing structural brain abnormalities, while Single-Photon Emission Computed Tomography (SPECT), Positron Emission Tomography (PET), and Cardiac Scintigraphy (123I-MIBG) help distinguish LBD from AD. Additionally, electroencephalography (EEG) helps identify slowed background activity, a characteristic often present in DLB.

D. Emerging Digital Biomarkers

Apart from handwriting analysis, which represents a promising non-invasive tool for the detection of LBD and other neurodegenerative disorders, other digital biomarkers have also been explored in recent decades. Speech analysis identifies subtle changes in voice and language patterns, which may indicate cognitive decline. Actigraphy, used for monitoring sleep disturbances (e.g., RBD) and daily activity levels, is also relevant for LBD. Gait analysis can track and analyse walking patterns for motor dysfunction assessment, which is commonly associated with neurodegenerative disorders. Integrating these digital biomarkers with handwriting analysis can enhance early diagnosis, improve disease monitoring, and contribute to developing personalized interventions, improving clinical research and patient care.

III. Handwriting characterization for LBD

Neurodegenerative disorders cause handwriting alterations because of motor and cognitive difficulties [3]. People affected by LBD often have difficulty with fine motor control, resulting in changes of handwriting velocity and altered stability of traits [4], [5]. The major signs of LBD in handwriting analysis include:

- Micrographia, shared with PD symptoms, with a progressive reduction in letter size;
- Irregular stroke pressure and velocity, resulting in reduced speed and fluidity in stroke execution, because of impaired motor control;
- Tremors and writing instability, linked to the extrapyramidal symptoms characteristic of LBD;
- Difficulties in spatial control and motor planning, attributable to cognitive deficits, such as visuospatial impairments;
- Increased pauses and longer task duration.

These signs affect handwriting, altering kinematic and dynamic handwriting parameters; thus, their analysis can support LBD diagnosis. Various tasks are considered to assess handwriting in LBD patients; in fact, they are widely used to characterize handwriting for other neurodegenerative disorders. Common evaluation tasks are as follows:

- Archimedean spiral drawing, a standard test for the diagnosis of dysgraphia in PD and evaluating the graphomotor difficulties. This task requires coordination and continued movement; thus, its execution can be compromised by the onset of LBD. Its analysis can highlight alterations in drawing speed, acceleration, fluidity, and precision, identified as discriminatory features.

- Sentence writing, which quantifies handwriting difficulties, highlighting those related to movement planning and coordination. In this task, the most informative features are the number of interruptions, the duration of in-air strokes, vertical speed on the surface, and vertical acceleration on the surface. Irregularities in these features are linked to cognitive deficits, fatigue, and loss of movement fluidity.
- Pentagon copying, used to assess cognitive decline and visuospatial abilities. This task requires visual perception and motor skills, revealing difficulties in drawing planning and execution, which are often more impaired in LBD, compared to AD [6].
- Trail Making Test (TMT), is a neuropsychological test to assess executive function, attention, and processing speed. It comprises two parts evaluating different cognitive abilities: TMT-A requires connecting numbers in sequence, while TMT-B requires alternating numbers and letters.
- Clock Drawing Test (CDT), used to assess visuospatial abilities, executive function, and the understanding of the concept of time.

These tasks are used in LBD diagnostics because each provides complementary details about cognitive and motor skills. The combined analysis of acquired parameters related to speed, pressure, pauses, and stroke quality improves the differentiation of LBD from other forms of dementia. It also helps with the monitoring of disease progression. Nowadays, handwriting analysis is facilitated by digital technologies, such as graphic tablets and sensor-equipped pens, opening the path to a non-invasive and promising method for the early diagnosis of LBD. Such tools enable the acquisition of detailed handwriting data, including key features such as speed, pressure, timing, pen trajectory, intra-stroke pauses, and in-air movements.

A. Handwriting Analysis and Artificial Intelligence

Once handwriting-related data are acquired, many analysis methods can be applied, along with the computation of other features. According to the literature, many approaches can be considered. The most conventional method takes into account parameters such as kinematic, spatial, and temporal parameters with the addition of pressure. Other methods involve the use of Fractional Derivatives [7], which is an advanced technique for parametrising kinematic data, taking into account temporal memory behaviour. Another approach considers Delta Parameters, a novel technique that extracts kinematic features by analyzing variations in speed and acceleration over time. Shannon entropy is often employed to assess the complexity and randomness of graphic gestures, providing a quantitative measure of variability in handwriting and drawing patterns. Additionally, the sigma-lognormal model [8] is used to model and parameterise movement, offering insights into a subject's motor control ability by describing the coordination and smoothness of their handwriting dynamics. These advanced techniques enhance the growing interest in handwriting analysis for a more detailed understanding of motor and cognitive impairments.

In recent decades, Artificial Intelligence (AI) techniques have been increasingly used in the medical context, particularly in analysing handwriting patterns to support the diagnosis of neurodegenerative disorders. AI can help identify LBD and distinguish it from other dementias. Apart from the classification, AI can also be exploited to monitor the disease progression, identify patients at risk, and detect early biomarkers of LBD. The combined evaluation of cognitive and motor functions using ML and DL improves diagnostic accuracy and contributes to the development of more effective monitoring and intervention strategies.

1) ML and DL Methods: Machine Learning (ML) and Deep Learning (DL) [9] models play a fundamental role in analysing handwriting data for the diagnosis of LBD. These models are used to classify patients, monitor and predict disease progression, and identify digital biomarkers. Classical ML algorithms, such as XGBoost, Support Vector Machines (SVM), and decision trees, have been exploited to distinguish between LBD, AD, and healthy controls, focusing on handwriting features for classification. Recent studies have provided concrete evidence supporting the use of handwriting and drawing analysis for LBD diagnosis. As shown in Table I, Galaz et al. [4] achieved 74% accuracy in detecting prodromal LBD using XGBoost as a classifier and features such as variability in spiral drawing and number of pen stops. Similarly, Yamada et al. [5] used Random Forest, reporting an AUC of 0.77 in distinguishing between AD, DLB, and controls using stylus-drawn tasks.

TABLE I. SUMMARY OF HANDWRITING-BASED STUDIES FOR LBD DIAGNOSIS

Study	Year	Data Type	Participants	Device	Task	ML Method	Result
Galaz et al. [4]	2022	Handwriting	39	Wacom Intuos 4 M	Archimedean spiral, sentence writing, pentagon copy	XGBoost	Accuracy: 74%
Yamada et al. [5]	2022	Handwriting/Drawing	123	Wacom Cintiq Pro 16	Sentence writing, pentagon copying, CDT, TMT tasks	Random Forest	AUC: 0.77

In more complex analyses, DL techniques, including Convolutional Neural Networks (CNNs) [10] and Recurrent Neural Networks (RNNs), have been applied, with CNNs excelling in spatial data analysis and RNNs being applied for sequential

handwriting trajectories. Regression models also contribute by predicting disease progression and correlating handwriting characteristics with clinical scores, providing a quantitative assessment of symptom severity. ML and DL approaches have demonstrated effectiveness in differential diagnosis, with classification accuracies ranging between 77% and 88%, highlighting the capability of their models to detect distinct patterns associated with different neurodegenerative conditions.

Digital biomarkers derived from handwriting (such as velocity, pressure, pauses, and fluidity) and advanced features (such as fractional derivatives and delta parameters) provide essential tools for diagnosing and monitoring LBD. These analyses enable the identification of specific motor and cognitive impairments, revealing reduced movement fluidity, increased pauses, and variability in drawing execution among LBD patients. Integrating ML and DL in handwriting analysis represents a promising non-invasive approach for LBD diagnosis and monitoring. AI may improve understanding of the disease and its early detection.

IV. Challenges and Future Directions

Research on LBD is facing multiple challenges: difficulties in early diagnosis, data heterogeneity, and limited amount of data; limited studies, particularly on disease progression; and the need for more interpretable AI models. The main obstacle remains the availability and the quality of data. Moreover, privacy concerns and data inconsistencies in the datasets limit the applicability of reliable and robust AI models. As stated in the previous sections, an early diagnosis of LBD is particularly challenging because its symptoms are similar to other neurodegenerative disorders. Also, few studies have evaluated AI models to track LBD progression.

Besides numerous challenges, many future directions can be considered. They include integrating multimodal data analysis, such as neuroimaging, handwriting, drawing, and speech data, to enhance our understanding of the disease and diagnostic accuracy. Transfer learning techniques could improve AI model robustness, whereas digital biomarkers, including handwriting dynamics and eye tracking, may aid in distinguishing LBD from similar conditions. Explainable AI methods, such as SHAP and LIME, can be applied more widely to improve transparency in decision-making. Additionally, AI has the potential to accelerate drug discovery by identifying pharmacological targets and optimizing clinical trials. Moving forward, a multidisciplinary approach combining AI, clinical research, and multimodal data analysis could significantly improve the diagnosis, monitoring, and management of LBD.

V. Conclusion

In conclusion, LBD presents significant diagnostic challenges because its symptoms overlap with other neurodegenerative disorders. Handwriting analysis has emerged as a promising, non-invasive, and economical tool for detecting motor and cognitive impairments associated with LBD. Digital tablets and AI-based methods offer precise kinematic measurements that can improve early detection and differentiation of LBD from similar conditions. However, the remaining challenges are data heterogeneity, limited datasets, and the need for explainable AI models. Future research should focus on integrating multimodal data, refining AI methodologies, and enhancing clinical interpretability to support accurate diagnosis and disease monitoring. A multidisciplinary approach that combines clinical expertise, neuroimaging, and advanced computational techniques holds great potential for advancing the early detection and management of LBD.

Acknowledgments

This study was partially funded by the Italian Ministry of Universities and Research (the PRIN 2022 LBDigital project, grant number 2022YXEP5T).

References

[1] A. Bougea, E. Efthymiopoulou, I. Spanou, and P. Zikos, "A novel machine learning algorithm predicts dementia with Lewy bodies versus Parkinson's disease dementia based on clinical and neuropsychological scores," *Geriatric Psychiatry and Neurology*, vol. 35, no. 3, pp. 317–320, May 2022.

[2] N. McCombe, A. Joshi, D. P. Finn, P. L. McClean *et al.*, "Distinguishing Lewy body dementia from Alzheimer's disease using machine learning on heterogeneous data: A feasibility study," in *2022 IEEE EMBC*, Jul. 2022, pp. 4929–4933.

[3] E. Onofri, M. Mercuri, G. Donato, and S. Ricci, "Cognitive fluctuations in connection to dysgraphia: A comparison of Alzheimer's disease with dementia Lewy bodies," *Clinical Interventions in Aging*, vol. 10, pp. 625–633, Mar. 2015.

[4] Z. Galaz, J. Mekyska, J. Mucha, V. Zvoncak, Z. Smekal, M. Faundez-Zanuy, L. Brabenec, I. Moravkova, and I. Rektorova, "Prodromal diagnosis of Lewy body diseases based on the assessment of graphomotor and handwriting difficulties," in *International Graphonomics Conference*. Springer, 2022, pp. 255–268.

[5] Y. Yamada, M. Kobayashi, K. Shinkawa, M. Nemoto, M. Ota, K. Nemoto, and T. Arai, "Characteristics of drawing process differentiate Alzheimer's disease and dementia with Lewy bodies," *Journal of Alzheimer's Disease*, vol. 90, no. 2, pp. 693–704, Jan. 2022.

[6] F. Cormack, D. Aarsland, C. Ballard, and M. J. Tovée, "Pentagon drawing and neuropsychological performance in dementia with Lewy bodies, Alzheimer's disease, Parkinson's disease and Parkinson's disease with dementia," *International Journal of Geriatric Psychiatry*, vol. 19, no. 4, pp. 371–377, 2004.

[7] J. Mucha, M. Gavenciak, J. Mekyska, M. Faundez-Zanuy, L. Brabenec, and I. Rektorova, "Evaluation of fractional calculus and delta parameters in prodromal diagnosis of dementia with Lewy bodies utilizing online handwriting," in *2024 32nd European Signal Processing Conference (EUSIPCO)*, 2024, pp. 1731–1735.

[8] T. D'Alessandro, C. Carmona-Duarte, C. De Stefano, M. Diaz, M. A. Ferrer, and F. Fontanella, "A machine learning approach to analyze the effects of Alzheimer's disease on handwriting through lognormal features," in *Graphonomics in Human Body Movement. Bridging Research and Practice from Motor Control to Handwriting Analysis and Recognition*, A. Parziale, M. Diaz, and F. Melo, Eds. Springer Nature Switzerland, 2023, pp. 103–121.

[9] N. D. Cilia, T. D'Alessandro, C. De Stefano, and F. Fontanella, "Offline handwriting image analysis to predict Alzheimer's disease via deep learning," in *2022 26th International Conference on Pattern Recognition (ICPR)*, 2022, pp. 2807–2813.

[10] S. Tasaki, N. Kim, T. Truty, A. Zhang, A. S. Buchman, M. Lamar, and D. A. Bennett, "Explainable deep learning approach for extracting cognitive features from hand-drawn images of intersecting pentagons," *NPJ Digital Medicine*, vol. 6, no. 1, p. 157, 2023.

Handwriting Measures as Predictors of Executive Functioning, Time Organization, and Self-Estimation of Students with and without Specific Learning Disabilities

Sara Rosenblum[1] and Kineret Sharfi[2]

1. Department of Occupational Therapy, Faculty of Social Welfare & Health Sciences and Laboratory for Complex Human Activity and Participation (CHAP), University of Haifa, Mount Carmel, Haifa 3498838 Israel. https://orcid.org/0000-0002-8666-0389. *2. Department of Occupational Therapy, Ben-Gurion University of the Negev, David Ben-Gurion Blvd. 1, Beer-Sheva, Israel.*
https://orcid.org/0009-0001-2227-8041

rosens@research.haifa.ac.il; sharfik@bgu.ac.il

***Abstract*—** Handwriting is a complex skill that integrates motor, cognitive, and perceptual processes, playing a crucial role in academic success and daily functioning. This study examines the interconnections between handwriting process measures, executive functions, time organization, and self-estimation in adults with and without Specific Learning Disabilities (SLD). A secondary analysis was conducted using a quasi-experimental case-control design with 110 participants (55 with SLD and 55 controls). Participants completed handwriting tasks on a digitizer alongside self-report questionnaires assessing executive functioning (BRIEF-A), time organization (TOPS), and self-esteem (SES).

Results indicated that while both groups adjusted pen tilt and stroke height between structured (paragraph copying) and unstructured (free writing) tasks, individuals with SLD exhibited greater variability in stroke height, suggesting deficits in fine motor control. Path analysis revealed that handwriting instability (pen tilt) related to deficits in inhibition, which in turn was associated with emotional responses related to time disorganization. Additionally, time organization and emotional regulation significantly predicted self-estimation, with individuals who managed time effectively reporting higher self-esteem. Notably, stronger working memory was paradoxically linked to lower self-estimation, possibly due to heightened self-awareness of cognitive difficulties.

The findings suggest that handwriting process measures such as pen tilt and stroke height may serve as potential biomarkers of deficits in fine motor control. These measures, associated with executive function deficits, particularly inhibition and working memory, may be related to broader implications on functional abilities, including academic performance and self-estimation. Thus, these results may be significant for individuals with SLD who may demonstrate decreased academic performance and self-estimation, compared to adults without SLD.

***Keywords*—** *Handwriting, Executive Functioning, Time Organization, Self-Estimation, Specific Learning Disabilities*

I. Introduction

Handwriting is a complex skill that integrates motor, cognitive, and perceptual processes, playing a fundamental role in academic success and daily functioning. It requires fine motor coordination, visual-motor integration, executive functioning, and the ability to plan, organize, and regulate written output efficiently [1]. For students with Specific Learning Disabilities (SLD), handwriting challenges are a prevalent issue, significantly affecting academic performance and self-esteem, [2]. Research highlights that students who frequently struggle with handwriting legibility, speed, automaticity, and consistency, are often frustrated and avoid writing tasks.

Studies have shown that students with SLD demonstrate greater variability in handwriting performance compared to their typically developing peers, particularly in free writing tasks, where cognitive and executive function demands are higher (3;4). This variability is attributed to deficits in fine motor coordination, visual-motor integration, sensory processing and executive functioning [2;4;5] and expressed in lack of movement precision, dysfluency and inconsistency in spatial and temporal measures [6]. Handwriting performance in SLD has been linked to broader executive functioning deficits, particularly in working memory, cognitive flexibility, inhibitory control, and planning [2]. It was found that students with SLD tend to show less variability during paragraph copy task compared to free writing, as the structured nature of the task reduces cognitive load [3]. However, even in copying tasks, their writing speed is significantly slower, and their letter formation remains inconsistent [2].

Beyond handwriting and executive function difficulties, time organization is another crucial challenge faced by students with SLD [5]. Time management, task prioritization, and adherence to structured routines are essential for academic success and self-regulation [7]. Research suggests that students with SLD struggle with time estimation, procrastination, and maintaining consistent

study habits, leading to increased frustration and academic underachievement [5]. Time disorganization is not only associated with academic struggles but also has significant emotional consequences. Research on adults with neurodevelopmental disabilities (NDDs) including those with SLD indicates that individuals with poor time management skills experience heightened stress, frustration, and emotional dysregulation[8]. Specifically, difficulties in planning, prioritizing, and following structured routines contribute to increased anxiety and negative emotional responses[8].

Frustration caused by inability to perform daily tasks as handwriting, and time disorganization among students with SLD may lead to academic anxiety and low self-esteem, which further exacerbates their learning difficulties[9]. Self-estimation, which refers to one's perception of their own abilities, can influence performance; if students perceive themselves as inefficient or experience repeated failures, they may develop low academic self-efficacy, further exacerbating writing difficulties [10].

Despite the well-documented link between executive functioning and handwriting difficulties, research examining the relationship between handwriting performance, EF abilities, broader time management skills and self-estimation in students with SLD remains limited. Given that handwriting performance, EF, and temporal organization are all interrelated in academic settings and may be tied with emotional health, a deeper understanding of these associations may promote theoretical and practical knowledge and lead to future development of focus intervention programs for students with SLD.

Figure 1 visually represents a theoretical model of these relationships.

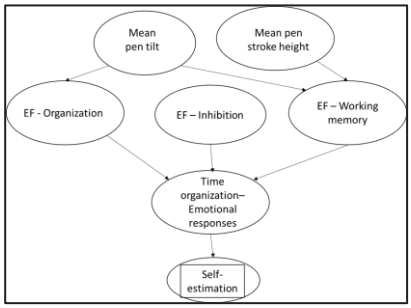


Fig. 1. Theoretical model for the relationships between handwriting performance measures, executive functions, time organization and self-estimation

Given these theoretical foundations, the present study investigates the interconnections between handwriting process measures, executive functions, time organization, and self-estimation in students with and without SLD. The following assumptions guide this study: Significant between and within group differences will be found for certain handwriting process measures across paragraph copying and free writing tasks. Certain handwriting process measures will predict executive functions (inhibition and working memory) and emotional responses due to deficient time organization abilities and all those variables will predict student's self-esteem.

II. Metodology

A. Procedure

This is a secondary analysis of quasi-experimental case-control design study which was approved at the Haifa University ethics committee (No. 170/10). Participants filled in a demographic questionnaire and the BRIEF-A, TOPS, and SES questionnaires (see details below). Each participant performed two writing tasks (a five lines paragraph copy and a free writing task- up to 10 lines) with an ink pen on a paper affixed to WACOM digitizer which was part of the ComPET-Computerized penmanship Evaluation Tool [11].

B. Participants

Based on the power analysis computation, a convenience sample of 110 adults ages 20-50 years participated in the study. Fifty-five of them diagnosed with SLD and 55 were controls see [12] for more details.

C. Instruments

Socio-demographic questionnaire. A self-report questionnaire which includes 36 questions was constructed for this study.

Behavioral Rating Inventory of Executive Functions – Adolescents/adults version (BRIEF-A) [13] is a . 75 items self-report questionnaire. aimed to examine behavioral manifestations of deficient executive functions (EF) abilities. Nine subscales are calculated: Inhibition; set-shifting; emotional control; self-monitoring; task initiation; working memory; planning/organization; task monitoring; and organization of materials.

Time Organization and Participation (TOPS) [14] is a 35 items self -report questionnaire aimed at examining difficulties in time organization while performing daily tasks. The questionnaire includes three parts with 5-point scales: at part A students are asked about their ability to perform daily activities in an appropriate pace, as expected by them B. they rate their performance in time

organization over the course of the day or within a certain period of time; and in part C. which was included at the model of the current study, they rate the frequency of emotional reactions which they feel as the result of a lack of success in performing activities at the pace expected of them.

The Self-Estimated Scale (SES) [15] is a widely used self-report measure designed to assess an individual's global self-esteem. The scale consists of 10 items, rated on a 4-point Likert scale ranging from 1 (strongly disagree) to 4 (strongly agree). Higher scores indicate greater self-esteem.

The ComPET- Computerized penmanship Evaluation Tool [11]. The computerized system enables the collection and analysis of spatial, temporal, and pressure handwriting data while the subject is writing on a paper affixed to a digitizer (an electronic tablet). The writing tasks were performed on A4 lined paper affixed to the surface of a WACOM Intuos 2 (model GD 0912-12X18) x-y digitizing tablet, using a wireless electronic pen with a pressure-sensitive tip (Model GP-110). Displacement, pressure and pen-tip angle are sampled at 100 Hz by means of a 1300 MHz Pentium (R) M laptop computer. The digitizer provides accurate temporal measures throughout the writing, both when the pen is touching the tablet (On-paper time) and when it is raised (In-air time). It also provides accurate spatial measures when the pen is touching the tablet and/or when it is lifted above the digitizer up to 6 mm [6;11].

D. Statistical Analysis

Descriptive statistics were calculated for socio-demographic characteristics. Preliminary t-test analysis indicated that the handwriting measures which differentiate between groups and indicated different appearance in each group in both handwriting tasks were mean pen tilt and mean pen stroke height. Due to high significant Pearson correlation between each of those two measures in both tasks, a mean score was computed to each measures of both tasks.

A repeated-measures General Linear Model (GLM) was conducted to examine differences in mean pen tilt (TLM) and mean stroke height (ShigM) across two writing tasks: paragraph copying and free writing, among students with and without SLD. The path analysis investigates the relationships between pen stroke height, pen tilt, working memory, inhibition, time organization -emotional reactions, and self-estimation.

III. Results

A. Socio-demographic characteristics

The group of adults with SLD (n=55) included 34.5% males and 64.5% females with mean age 29.58 (SD = 6.4) and the control group (n=55) included 23.6% males and 76.4% females with mean age 31.18 (SD = 6.4). See 12 for further details

B. Handwriting process measures

TABLE 1: HANDWRITING PROCESS MEASURES ACROSS TASKS, BETWEEN AND WITHIN GROUPS COMPARISON

Measure	**Task**	**Group**	**M (SD)**	**Group Effect (*F*, *p*, η^2)**	Task Effect (*F*, *p*, η^2)	Interaction (*F*, *p*, η^2)
TLM (Mean pen tilt)	**Paragraph copying**	Control	60.05 (5.36)	2.44, 0.12, 0.02	41.53 <0.001 0.28	0.046, 0.83, 0.00
		SLD	58.37 (5.51)			
	Free writing	Control	61.82 (6.21)			
		SLD	60.25 (5.52)			
ShigM (mean stroke height)	**Paragraph copying**	Control	0.24 (.05)	9.35, 0.003, 0.080	9.19, 0.003, 0.78	0.00, 0.95, 0.00
		SLD	0.28 (0.07)			
	Free writing	Control	0.23 (0.05)			
		SLD	0.27 (0.07)			

As presented in table 1, a significant main effect of task ($p < 0.001$, $\eta^2 = 0.28$) indicates that pen tilt differed between copying and free writing while no significant overall difference was fond between SLD and control groups ($p = 0.12$, $\eta^2 = 0.022$).

A significant main effect of task ($p = 0.003$, $\eta^2 = 0.078$) and group effect ($p = 0.003$, $\eta^2 = 0.080$) which was found for mean pen stroke height indicate that students with SLD exhibited greater stroke height variability than controls.

C. Path analysis results

The path analysis investigates the relationships between pen stroke height, pen tilt, working memory, inhibition, time organization - emotional reactions, and self-estimation. Results indicated good model fit (CMIN/DF (χ^2/df) = 1.009 ($p = 0.422$) GFI = 0.979, AGFI = 0.937, CFI = 0.999, TLI = 0.999, IFI = 1.000 RMSEA = 0.09 (90% CI: 0.000 - 0.118, $p = 0.059$). Meaning that the model fits the data well, meaning the proposed relationships among the variables are statistically supported.

1) Regression Paths (Direct Effects)

A. Predicting inhibition (EFinhibit) and working memory EFTWM) by handwriting process measures (pen tilt-TLM12, SHIGM12): 1. Pen tilt → Inhibition: β = -0.134, p = 0.041 2. Pen stroke height → Working Memory: β = 0.171, p = 0.009

B. Predicting emotional reactions due to deficient time organization (TOPSCM) by inhibition and working memory : 1. Inhibition Emotional Reactions Time Organization β = -0.497, p < 0.001, 2. Working Memory → Emotional Reactions Time Organization: β = -0.118, p = 0.279 (not significant)

C. Predicting Self-Estimation (SESMean) by Emotional Reactions Time Organization, inhibition working memory, and handwriting process measures (pen tilt): 1. Time Organization - Emotional Reactions → Self-Estimation: β = 0.424, p < 0.001, 2. Working Memory (EFTWM) → Self-Estimation: β = -0.251, p = 0.00 3. Pen Tilt → Self-Estimation: β = 0.196, p = 0.009

2) Indirect Effects (Mediated Relationships)

1. Pen Tilt → Inhibition → Time Organization → Self-Estimation: Standardized indirect effect = -0.052, p = 0.014. 2. Working Memory → Time Organization → Self-Estimation: Standardized indirect effect = -0.050, p = 0.029, 3. Inhibition → Time Organization → Self-Estimation: Standardized indirect effect = -0.211, p = 0.002

IV. Discussion

By analyzing handwriting process measures alongside executive functioning, time organization, and self-estimation, this study aims to broaden the understanding of handwriting performance features of students with SLD and their relationships with functional abilities at the above mentioned domains. The results of the current study add to previous results and suggest that while both groups adjust their pen tilt and stroke height depending on the writing task, the SLD group exhibits greater overall stroke height variability, possibly reflecting difficulties in fine motor control and writing stability. The lack of interaction effects suggests that task demands (copying vs. free writing) influence both groups similarly, yet individuals with SLD consistently show greater inconsistencies in stroke height across writing conditions.

The path analysis exhibited interesting results. Greater motor instability as exhibited by pen tilt may reflect underlying deficits in motor planning and executive function, affecting inhibitory control. On the other hand, the fact that higher stroke height variability is associated with better working memory may demonstrate that higher cognitive flexibility in letters size allowing for better working memory performance.

Inhibition ability is significantly tied not only with pen tilt but also with emotional reactions due to deficient time organization ability as decreased inhibition may contribute to impulsivity and emotional dysregulation, making it harder to manage time effectively.

Results further indicate that better time organization and emotional regulation lead to higher self-estimation. Individuals who can effectively manage their time and emotions may have greater self-confidence and self-awareness leading to higher self- esteem.

A surprising result indicates that better working memory is associated with lower self-estimation. Students with stronger working memory might be more aware of their own cognitive limitations, leading to a more critical self-view.

As expected that better time organization and emotional regulation lead to higher self-estimation, individuals who can effectively manage their time and emotions may have greater self-confidence and self-awareness.

In sum, handwriting process measures of handwriting activity as pen tilt and stroke height may be specific biomarkers of human's executive functions as inhibition and working memory and of functional abilities of students with SLD and tied to their self-estimation. A deep understanding of these measures and their associations with functional abilities may lead to development of future focused intervention programs for students with SLD.

References

[1] K. P. Feder, and A. Majnemer, "Handwriting development, competency, and intervention," Dev Med Child Neurol. vol. 49, pp. 312–317, 2007.

[2] A. L. Barnett, M. Prunty, ans S. Rosenblum, "Development of handwriting speed and legibility in children with and without dyslexia over the course of the primary school years," Read Writ. vol. 31, pp. 1063–1084, 2008.

[3] M. Prunty, A. L. Barnett, K. Wilmut, and M. Plumb, "Visual perceptual and handwriting skills in children with Developmental Coordination Disorder," Hum Mov Sci. vol. 49, pp. 54–65, 2016.

[4] K. R. Harris, & S. Graham, "Self-regulated strategy development: Theoretical bases, critical instructional elements, and future research," in Design principles for teaching effective writing, R. Fidalgo Redondo, K. Harris, and M. Braaksma, Eds. Brill, 2017, pp. 119–151.

[5] E. Sumner, V. Connelly, and A. L. Barnett, "Children with dyslexia are slow writers because they pause more often and not because they are slow at handwriting execution," Read Writ. vol. 26, pp. 991–1008, 2013.

[6] S. Rosenblum, D. Chevion & P. L. Weiss, "Using data visualization and signal processing to characterize the handwriting process," Pediatr Rehabil. vol. 9, pp. 404-417, 2006, DOI: 10.1080/13638490600667964

[7] R. A. Barkley, Executive Functions: What they are, how they Work, and why they Evolved. Guilford Press, 2012.

[8] S. Rosenblum, K. Sharfi, O. C. Elimelech, S. Regev, and N. Grinblat, “Predicting emotional consequences of disorganization among adults with neurodevelopmental disorders: A structural equation model,” Disabil Rehabil. pp. 1–11, 2014. https://doi.org/10.1080/09638288.2024.2385733

[9] S. S. Zentall, “Executive dysfunction and academic performance in students with learning disabilities,” J Educ Psychol. vol. 110, pp. 725–740, 2018

[10] . J. Cleary, B. Velardi, and B. Schnaidman, “Effects of the Self-Regulation Empowerment Program (SREP) on middle school students' strategic skills, self-efficacy, and mathematics achievement,” J Sch Psychol. vol. 64, pp. 28-42, 2017.

[11] S. Rosenblum, S. Parush, and P. L. Weiss, “Computerized temporal handwriting characteristics of proficient and non-proficient handwriters,” Am J Occup Ther. vol. 57, pp. 129–138, 2003.

[12] K. Sharfi, and S. Rosenblum, “Sensory Modulation and Sleep Quality amongAdults with Learning Disabilities: A Quasi-Experimental Case-Control Design,” PloS ONE, vol. 10, e0115518, 2015.

[13] R. M. Roth, P. K. Isquith, and G. A. Gioia, “Behavior Rating Inventory of Executive Function: Adult version (BRIEF-A),” Psychological Assessment Resources, 2005.

[14] S. Rosenblum, “The development and standardization of the Time Organization and Participation Scale (TOPS) for children,” Am J Occup Ther. vol. 66, e42–e50, 2012. https://doi.org/10.1111/j.1365-2214.2006.00687.x

[15] M. Rosenberg, Society and the adolescent self-image. Princeton: Princeton University Press, 1965.

Leveraging Grammatical Evolution for Interpretability in Computer-Assisted Diagnosis of Alzheimer's Disease

Amos Battista
NClab, DIEM, University of Salerno
amosbattistav@gmail.com

Antonio Della Cioppa
NClab, DIEM, University of Salerno
adellacioppa@unisa.it

Angelo Marcelli
NClab, DIEM, University of Salerno
amarcelli@unisa.it

***Abstract*— This work focuses on the early, non-invasive diagnosis of Alzheimer's Disease (AD) through handwriting analysis. Although advanced algorithms can yield high accuracy, medical professionals require transparent reasoning to trust automated decisions. This research therefore aims to identify AD symptoms while preserving interpretability In the literature, diagnosing AD often relies on expensive or invasive procedures, limiting scalability and patient comfort. Handwriting analysis offers a simple, cost-effective way to detect subtle neurodegenerative signs. Yet, many current models are opaque black-boxes with little insight into their workings. To address this issue, Grammatical Evolution is employed to evolve Decision Trees inherently preferred for their straightforward, interpretable rule-based structure—thus avoiding the pitfalls of post-hoc explanations. Experiments conducted on the DARWIN dataset confirm that state-of-the-art performance can be achieved without compromising transparency, offering a genuinely interpretable solution.

Keywords*— *Neurodegenerative diseases, Interpretable Artificial Intelligence, Grammatical Evolution, Decision Trees.

I. Introduction

This work aims at creating interpretable AI models for handwriting analysis-based early Alzheimer's disease diagnosis. The central purpose is to evaluate whether the benefits of interpretability could be achieved without a significant trade-off in classification performance, particularly in a critical domain such as medical diagnostics, where transparency is paramount. The strength of the proposed approach relies on the use of Grammatical Evolution (GE), using a specialized grammar that evolves intrinsically interpretable Decision Trees models (DTs), thus eliminating the need for post-hoc explanation methods typically required for black-box models. Novelty comes from using GE to obtain DTs tailored for interpretability when applied to Alzheimer's diagnosis. The performance of the approach has been evaluated on the DARWIN dataset, consisting of handwriting samples from Alzheimer's patients and healthy controls. The results show that the DTs produced by the GE achieve classification performance comparable to traditional machine learning methods, while exhibiting reduced complexity, thus suggesting that the GE approach proposed in this paper is the most promising, offering the best trade-off between performance and interpretability. Finally, an adaptive ensemble approach further improved the classification performance, yielding the highest overall accuracy of all setups.

II. The Grammatical Evolution Approach

GE [1] is an advanced paradigm that extends the principles of genetic evolution to higher levels of abstraction and complexity. While it shares foundational concepts with Genetic Programming (GP), GE distinguishes itself through unique mechanisms and outcomes, offering innovative solutions in theoretical and applied domains. Similarly to GP, in GE a genome is represented as a string of integers, which are transformed into a set of syntactic rules constituting the phenotype, i,e., the solution to the problem that the computational system aims to solve. The core of GE lies in the grammar that governs the generation of solutions. Unlike other evolutionary algorithms, GE is based on formal grammars defined using Backus-Naur Form (BNF). As a consequence, the design of the grammar is pivotal in steering the evolutionary process toward generating models that are not only accurate but also interpretable.

A. Integration of Decision Trees into GE

DTs are inherently interpretable models that offer transparent pathways from input features to output predictions. Each node in a DT represents a decision based on a feature, and every path from the root to a leaf node signifies a specific classification rule. Figure 1 reports the BNF representation of the grammar used. By using the above grammar to produce inherently interpretable models, post-hoc explanations are not needed. This approach is in line with the principles of interpretable AI [2], which ensure that the models are transparent, trustworthy, and efficient [3], [4].

III. Experimental Findings

The experimental framework adopted aims to fill the gap between theoretical and practical investigation by showing how GE can respond to the dual demands of accuracy and interpretability. To provide the final decision, we adopt an ensemble approach that aggregates the predictions provided by the best DTs from all 25 task-specific experiments by means of a majority vote mechanism. The rationale behind the ensemble comes from the complementary nature of the tasks: each task encapsulates

```
<p> ::= <defp>{::}result = p()
<defp> ::= def p():{:<if-else>{::}return <TF>:}
<if-else> ::= if <f> <op> <num>:{:<if-else>:} | if <f> <op> <num>:{:<if-else>:}else:{:<if-else>:} | return <TF>
<f> ::= x[0] | x[1] | ... | x[17]
<op> ::= ≥ | ≤
<num> ::= <int> | <dec>
<int> ::= <digit_non_zero> <digits_opt> | 0
<digits_opt> ::= <digit> GE_RANGE:9
<dec> ::= <int>.<digit> | <int>.<digit><digit> | 0.<digit> | 0.<digit><digit>
<digit> ::= 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9
<digit_non_zero> ::= 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9
<TF>::= True | False
```

Fig. 1: The DT grammar embedded within GE.

unique aspects of motor and cognitive functions, and thus their integration compensates for the limitations inherent in the individual tasks.

The approach adopted was implemented as follows:

- **Baseline experiment:** traditional DTs were first used to establish a performance benchmark for each of the 25 tasks as well as for their MV ensemble.
- **GE experiments:** these experiments aimed at: i) maximizing accuracy to improve predictive reliability; ii) optimizing the F1-score, which is essential in clinical contexts for balancing precision and recall; iii) applying complexity penalties to simplify tree structures, thereby enhancing interpretability without sacrificing performance.
- **Adaptive ensemble experiment:** the experiment brought in Majority Vote (MV) ensemble strategies that uses of an ensemble of decision trees for each task, leveraging the best-performing models in the final population as identified by the evolutionary process. Unlike earlier setups where a single DT per task was used, this experiment allows to dynamically determine the optimal number of trees required, depending on their effectiveness during the training. More specifically, unlike previous setups where only one majority vote was conducted across task-specific best predictions, this new framework added an internal majority vote for each task. Within each task, the retained ensemble of decision trees cast votes to determine the task-level classification. These task-specific decisions are then aggregated across tasks through the second majority voting step, producing the final diagnosis. This dual voting system amplifies the robustness of the diagnostic process by applying majority voting to the predictions within each task (*Intra-Task Voting*), the system mitigates the risk of individual tree biases, ensuring more reliable task-level decisions. The final majority vote across all task-level predictions (*Inter-Task Voting*) provides a cohesive and balanced diagnostic output. However, with multiple DTs influencing each task-level prediction and an additional layer of aggregation across tasks, tracing the rationale behind the final diagnosis became increasingly opaque. For clinical applications, where transparency and trust are vital, this layered decision-making process introduces a barrier to usability, highlighting the ongoing trade-off between accuracy and usability in machine learning systems designed for healthcare. To mitigate this problem, an adaptive strategy has been adopted. During the training, only the trees in the population that demonstrably improves the ensamble performance on the training data are retained. As a result, this mechanism results in ensembles comprising few models.

Each experimental phase builds iteratively, incorporating insights from its predecessor to refine both methodology and outcomes. The following sections systematically navigate these experimental phases in order to comprehensively investigate the potential of GE and form a basis for future developments in interpretable AI and their integration into healthcare.

A. DARWIN dataset

The DARWIN dataset [5] is a valuable resource for early detection research in Alzheimer's disease using handwriting analysis. It consists of 174 participants who provided handwriting samples, of which 89 were diagnosed with Alzheimer's disease and 85 were healthy [5]. Each participant performed 25 writing tasks, carefully designed to assess various cognitive and motor skills related to handwriting, grouped into 4 main categories, each focusing on distinct aspects of fine motor control and cognitive functions. Each task produced a rich dataset with 18 kinetic and dynamic features, so as we have 25 different data sets. Each data set was firstly normalizied in the range [0,1] and, then, was randomly partitioned into 25 distinct training and testing splits (the occurrence of 25 tasks in the dataset and the 25 experimental splits is entirely coincidental). This methodological approach was designed to evaluate the performance across diverse configurations, thus avoiding dependence

on a single train-test division. Each split provided a unique combination of training and testing data, simulating different real-world scenarios and yielding a nuanced understanding of the generalization capabilities of the model.

B. Results

1) Baseline Experiment - Decision Tree Benchmark: a standard DT classifier was applied to the DARWIN dataset. using Gini impurity and entropy as criteria for tree splitting. A grid search approach was used, to fine-tune critical hyper-parameters. Accuracy across tasks typically ranged between 55% and 65%, with some variability depending on the task. Precision, recall, and F1 scores confirmed this pattern, revealing that certain tasks offered higher predictive utility due to their stronger correlation with Alzheimer's impairments. The MV mechanism improved the overall average accuracy to about 77%, reflecting the benefit of aggregating predictions from multiple tasks. However, the ensemble's accuracy highlighted the limitations of relying on single-task models for robust diagnostics. One of the main challenges was the complexity of the DTs. Many trees were very deep, up to more than 15 levels deep, with a huge number of nodes and leaves. These characteristics, while indicative of the models' efforts to capture intricate patterns in the data, severely hindered interpretability.

2) First Experiment – GE Accuracy Maximization: this experiment was a fundamental step in transitioning from traditional DT methodologies to the GE approach. While improving classification performance was a secondary aim, the primary goal was to investigate how GE could autonomously balance feature selection, tree complexity, and interpretability. GE was equipped with a population of 100 randomly generated genomes, each encoding a DT, and at the end of the evolution, the MV aggregated predictions across all tasks, reinforcing the robustness of the diagnostic system.

Accuracy across single tasks typically ranged between 62% and 72%, achieving a clear improvement over the previous result. The ensemble accuracy, gained by the MV mechanism, reached an average of 82%. This underscored the efficiency of the evolutionary approach in exploring a larger search space and identifying more effective DT structures. The significant steps of this experiment involve the dynamic feature selection provided by the evolutionary framework: many features were used in about 60-70% of the tasks, which suggests a fair and even distribution of feature importance. Moreover, the average number of features used per task dropped from 18 to 13. This means that fewer features than the baseline experiment are used for each task. This decrease goes beyond simple optimization—it acts as a key step in reducing the width of the DTs. However, GE-generated DTs were rather complex, partially attributable to the fitness function, which solely optimized accuracy. While this focus improved predictive performance, it inadvertently encouraged overfitting in some cases, leading to intricate decision paths that reduced the practical utility of the trees. Moreover, a closer examination of the DTs evolved by GE reveals some critical aspects. Many of the generated trees extended to depths exceeding 10 levels, proving to their ability to capture intricate patterns but also a significant barrier to their practical interpretability. The average tree contained over 50 nodes and approximately 20 leaves, representing the complexity of the decision pathways formed to achieve high performance. Despite the introduction of feature selection mechanisms, most trees relied on 12 to 14 features, demonstrating a partial reduction in dimensionality but highlighting the need for further streamlining. These factors collectively underscore the difficulty of balancing accuracy with interpretability.

3) Second Experiment – GE F1-score Maximization: here, it was considered important to study whether the optimization of F1-score could return more balanced DTs. The results show that, while the overall performance remained comparable to the previous experiment, slight but important differences emerged in key metrics. Recall increased from 75% to 78%. By focusing on F1-score, the algorithm prioritized identifying true positives, ensuring fewer patients were overlooked by the model. Precision decreased marginally from 91% to 88%. This drop was expected, as the system compensated for higher recall by slightly over-classifying healthy individuals as Alzheimer's cases. Clinically, this trade-off may be acceptable, as the consequences of missing a diagnosis typically outweigh the burden of further testing for false positives. However, the unchanged accuracy suggests that this balancing act had little overall impact on the system's ability to distinguish between healthy and Alzheimer's patients. The shift to F1-score optimization introduced minimal impact on the complexity of the evolved DTs. Many trees exhibited depths exceeding 15 levels, reflecting the system's continued reliance on intricate decision paths to achieve high performance. On average, the trees contained over 55 nodes and 22 leaves, like the accuracy-focused experiment. This persistence underscores the challenge of controlling tree growth when fitness functions do not explicitly penalize complexity. The number of features used per tree remained consistent, averaging 12-14 features.

The experiment underscored the limitations of relying solely on predictive metrics for fitness evaluation. The DTs remained large and intricate, limiting their practical utility in clinical settings where simplicity and transparency are paramount.

4) Third Experiment – GE Balancing Accuracy and Complexity: this experiment introduced a new optimization approach that strikes a balance between classification accuracy and tree complexity, aiming to address the trade-offs that had previously hindered practical interpretability. This dual-objective function was mathematically expressed as: $\phi = \mathrm{A} - p \cdot \frac{C}{n_{\max}}$, where A is the accuracy, C is the # of nodes, $n_{\max}$ is the maximum # of allowed nodes, and p is a penalty factor set to 0.05. The third experiment produced significant results, proving that the potential of this approach to produce models was not only effective but also interpretable and clinically feasible. The use of a penalty for model complexity significantly influenced the structure of the evolved DTs, resulting in models more concise without compromising diagnostic accuracy. A detailed

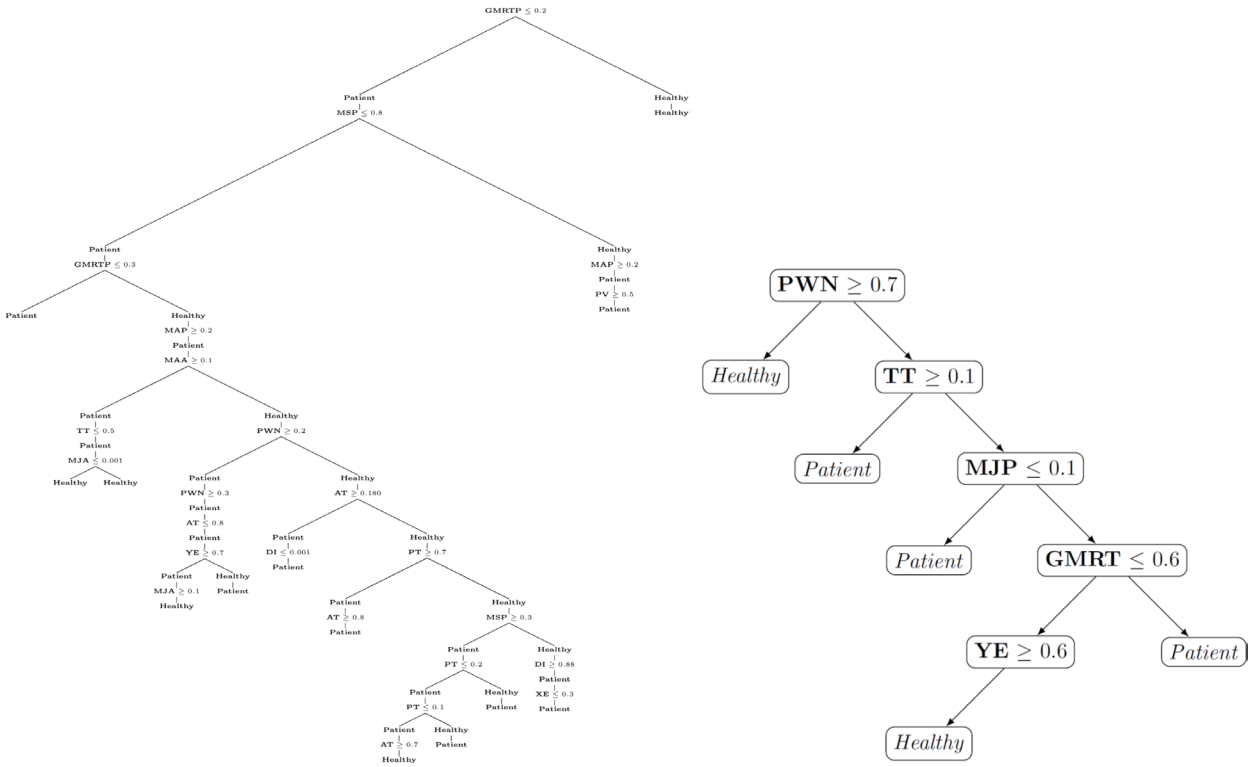
examination of the classification metrics revealed subtle yet meaningful shifts. The average precision remained high at 91%, demonstrating the model's ability to minimize false positives effectively. This metric's consistency suggests that the penalty for complexity did not adversely impact the model's capacity to correctly identify healthy individuals. Recall stabilized at 78%, maintaining the gains observed early. The overall accuracy of 84% marked a good improvement, which is particularly noteworthy because it was achieved alongside a significant reduction in tree complexity. Indeed, the average number of features per DT dropped significantly, from 12-14 in first experiment to approximately 4-5, underscoreing the system's ability to prioritize diagnostically significant features. The explicit penalization mechanism introduced in the fitness function played a central role toward interpretability, targeting both the size and structure of the trees: The average normalized complexity, calculated by dividing the number of nodes by the maximum threshold of 40, dropped by nearly 50% compared to earlier experiments. By explicitly discouraging overly complex trees, the system systematically evolved more compact and streamlined decision paths without sacrificing accuracy. This complexity reduction was evident in both the depth and size of the DTs. The average tree depth decreased from over 15 levels in the earlier experiments to fewer than 8 levels in this experiment. This marked reduction represents a paradigm shift, as simpler models are far more accessible for visualization, explanation, and clinical validation. The average number of nodes per tree dropped to 8, an absolute contrast to the 50+ nodes observed in the previous setups. The average number of leaves per tree significantly decreased from 20 to 3, indicating a substantial simplification of decision paths and a more efficient representation of the classification logic. Beyond the numerical metrics, this experiment paved the way towards developing models that would be understandable and, at the same time, accurate. The average number of nodes dropped significantly, bringing the trees closer to a level that clinicians could feasibly analyze and trust.

5) Fourth Experiment – GE Adaptive Ensemble Approach: the fourth experiment marked a significant shift from previous ones, prioritizing accuracy above all else. This approach involved the use of an ensemble of DTs for each task, leveraging the best-performing models as identified by the evolutionary process. Unlike earlier setups where a single tree per task was used, this experiment allowed the system to dynamically determine the optimal number of trees required, depending on their effectiveness during training. To generate a larger pool of high-quality models, the population size was increased from 100 to 500 genomes, facilitating the discovery of diverse and effective tree structures. The increase in population size ensured that the top candidates represented a genuinely diverse and high-performing subset. A critical enhancement in this experiment was the introduction of a dual majority voting mechanism, where only one MV was conducted across task-specific predictions, adding an internal MV for each task. Within each task, the retained ensemble of DTs cast votes to determine the task-level classification. These task-specific decisions were then aggregated across tasks through the second majority voting step, producing the final diagnosis. This dual voting system amplified the robustness of the diagnostic process, ensuring more reliable task-level decisions. The final MV across all task-level predictions provided a cohesive and balanced diagnostic output. The results highlighted the advantages of this approach, demonstrating the system's ability to adapt dynamically, while ensuring precision without unnecessary computational burden. However, this complexity came at the expense of interpretability. With multiple DTs influencing each task-level prediction and an additional layer of aggregation across tasks, tracing the rationale behind the final diagnosis became increasingly opaque. The model's enhanced performance, while impressive, must be weighed against its diminished interpretability, highlighting the ongoing trade-off between accuracy and usability. A key innovation was the adaptive selection of DTs for each task. While the system initially identified many trees based on the same fitness function used in the previous experiment, not all these models were necessarily included in the final ensemble. During training, only those trees that demonstrably improved performance on the training dataset were retained. This mechanism often resulted in ensembles comprising few models, with some tasks requiring as few as 4-8 trees. The number of trees per task varied significantly. However, this experiment achieved record classification performance, with the highest average accuracy across all setups at an impressive 88%, compared to the 84% recorded in the third experiment. This substantial gain underscores the transformative potential of ensemble strategies, which leverage the collective strengths of multiple models to extract and reinforce reliable patterns within the data. An 88% diagnostic accuracy represents a notable advancement in early Alzheimer's detection, providing clinicians with a reliable tool to identify patients requiring further evaluation. Precision stabilized at an impressive 95%, reflecting the system's effectiveness in minimizing false positives. However, recall was slightly lower at 82%, indicating that while precision was high, the system missed some Alzheimer's cases. The ensemble approach achieved unprecedented accuracy, showcasing the system's theoretical potential when interpretability is deprioritized. However, the added complexity of multiple DTs per task and the layered majority voting system compromised the interpretability. While the system's performance and robust decision-making are impressive technical achievements, they come with significant challenges in transparency and usability. Therefore, this experiment serves as a benchmark for performance optimization rather than a practical diagnostic tool. These results confirm the effectiveness of the evolutionary approach in producing interpretable yet high-performing models. Compared to traditional DTs, the introduction of multi-objective optimization (accuracy vs. complexity), the focus on the F1-score, and the reduction in nodes, depth, and feature usage allowed for a desirable balance between diagnostic accuracy and model simplicity. The addition of the adaptive ensemble approach further enhanced performance, achieving up to 88% accuracy, but at the cost of reduced transparency, underscoring the trade-off between accuracy and interpretability.

Table I summarizes the key results from the experiments, aggregating accuracy, F1-score, and tree complexity metrics,

	Avg Task Accuracy	Avg # of Nodes	Avg DT Depth	Avg # of Features	Avg MV Accuracy	Avg Precision	Avg Recall
Decision Tree	0.59 ± 0.02	54.12 ± 15.20	18.20 ± 7.20	18.00 ± 0.00	0.77 ± 0.06	0.74 ± 0.06	0.88 ± 0.08
GE - Accuracy	0.67 ± 0.08	58.88 ± 29.20	16.89 ± 6.35	13.02 ± 0.70	0.82 ± 0.06	0.75 ± 0.08	0.91 ± 0.08
GE - F1 Score	0.66 ± 0.08	61.10 ± 27.10	17.04 ± 6.02	13.14 ± 0.49	0.82 ± 0.05	0.78 ± 0.10	0.88 ± 0.05
GE - Accuracy/Complexity	0.67 ± 0.07	7.86 ± 3.24	4.94 ± 1.50	4.03 ± 0.53	0.84 ± 0.07	0.78 ± 0.09	0.91 ± 0.08
GE - Adaptive ensamble	0.70 ± 0.02	9.32 ± 2.13	5.34 ± 0.49	7.20 ± 0.93	0.88 ± 0.05	0.82 ± 0.07	0.95 ± 0.06

TABLE I: Summary of key performance metrics across all experiments, including Average Task Accuracy, Average # of Nodes, Average DT Depth, Average # of Features, Average Majority Voting Accuracy and Average Precision and Recall.



(a) Best DT obtained form the baseline experiment (DT approach). (b) Best DT obtained by the GE form third experiment.

Fig. 2: Comparison between the DT obtained by the baseline experiment and the best one obtained by GE on the 24-th task.

while Figure 2 shows the best DT obtained form the baseline experiment (left pane) and the best DT from the balancing accuracy and complexity one (left pane) on the 24-th task. The reduction in complexity is evident in both the depth and the size of the GE-generated DT. Similar complexity reductions occur for all the other tasks.

IV. Conclusions

The aim of this work is to leverage GE to generate interpretable decision tree models, specifically tailored for applications requiring transparency and trustworthiness, such as medical diagnostics. By designing a grammar that encapsulates the structure and logic of DTs, GE can evolve models that are both accurate and accessible for human interpretation. This methodology contributes to the growing field of Interpretable AI by providing a practical solution for generating transparent models. It bridges the gap between the need for high-performing models and the requirement for interpretability in critical applications.

References

[1] M. O'Neill and C. Ryan, "Grammatical evolution," *IEEE Transactions on Evolutionary Computation*, vol. 5, no. 4, pp. 349–358, 2001.
[2] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, p. 206–215, 2019.
[3] A. Parziale, R. Senatore, A. Della Cioppa, and A. Marcelli, "Cartesian genetic programming for diagnosis of parkinson disease through handwriting analysis: Performance vs. interpretability issues," *Artificial Intelligence in Medicine*, vol. 111, p. 101984, 2021.
[4] A. Della Cioppa, I. De Falco, T. Koutny, U. Scafuri, M. Ubl, and E. Tarantino, "Reducing high-risk glucose forecasting errors by evolving interpretable models for type 1 diabetes," *Applied Soft Computing*, vol. 134, p. 110012, 2023.
[5] N. Cilia *et al.*, "Diagnosing alzheimer's disease from on-line handwriting: A novel dataset and performance benchmarking," *Engineering Applications of Artificial Intelligence*, vol. 111, p. 104822, 2022.

Multimodal Handwriting Analysis in Alzheimer's Disease Detection through Bayesian Networks

Emanuele Nardone, Tiziana D'Alessandro, Claudio De Stefano, Francesco Fontanella, Alessandra Scotto di Freca
Università degli Studi di Cassino e del Lazio Meridionale Via Gaetano di Biasio, 43, 03043 Cassino FR
{emanuele.nardone, tiziana.dalessandro, destefano, fontanella, a.scotto}@unicas.it

***Abstract*— Alzheimer's disease is recognised as the most widespread neurodegenerative disorder worldwide. Previous research introduced a protocol that employs handwriting tasks as a potential diagnostic tool for predicting Alzheimer's disease. Our study extends this protocol by analysing both online handwriting acquisitions, capturing static and dynamic features, and offline acquisitions. In contrast, handwriting images are processed using a CNN as a feature extractor. Our proposed approach uses a Bayesian Network to combine the multimodal handwriting data and predict the onset of Alzheimer's disease. This work aligns with the growing trend in healthcare of using multimodal machine learning methods to support clinical decision-making. It shows an important improvement in detecting Alzheimer's disease onset with respect to both online and offline acquisitions.**

Keywords*— *Medical Expert Systems, Cognitive Impairments, Bayesian Networks, Multimodal Handwriting Analysis, Classifier Combination.

I. Introduction

Healthcare machine learning has traditionally focused on single-modal data, but recent efforts are increasingly incorporating multimodal information to mimic clinical decision-making better. Alzheimer's disease (AD), the most common form of dementia, affects multiple cognitive functions and is typically diagnosed using a combination of physical exams, medical history, cognitive assessments, neurological tests, and neuroimaging [1]. Early detection of AD is crucial as it provides a critical window for therapies that can slow disease progression, reduce care needs, and extend periods of healthy living. Notably, emerging research in handwriting analysis offers a less invasive method for early AD detection by capturing subtle changes in cognitive processes and motor control associated with the disease [2], [3]. In a previous study [4], a protocol with 25 handwriting tasks was introduced to capture the dynamics of individuals with cognitive impairments using a graphic tablet that recorded both pen movements and ink traces. Subsequent research applied feature selection methods [5] to analyse in-air and on-paper movements, while CNNs were employed to extract features from handwriting images [6], capturing variations in shape, size, and thickness. Despite these advances, previous studies either treated tasks as isolated information sources or merged them to increase sample size, without exploring the potential of multimodal data fusion. Our recent work investigated online handwriting features through a Bayesian Network (BN) for AD detection [7]. The present study extends this approach by including offline handwriting features through CNN analysis, providing a more extensive multimodal framework. However, multimodal diagnostic models have shown an average improvement of 6.4% in predictive accuracy compared to unimodal approaches, although their implementation still presents challenges regarding scalability and data integration [8]. The present work aims to use multimodality by integrating online and offline handwriting features through a BN. A BN is a probabilistic graphical model widely used in the medical field. In [9], BNs were used to model the dependencies between symptoms and diseases to suggest the most probable diagnosis. In [10] BNs integrate multimodal sensory data for a robust understanding of the environment, and in [11], they uncover relationships among neuroimaging markers and cognitive performance indicators in preterm and full-term subjects. A BN depicts a set of random variables using a directed acyclic graph (DAG). In the DAG, each node represents a variable, which, in our case, is the prediction for online or offline tasks or the AD label. The BN models conditional dependencies among variables, and the edges between nodes indicate the dependency, whose values are quantified in conditional probability tables. By focusing on the AD label, only a subset of task predictions remain connected to the AD label in the DAG; this subset is called the Markov Blanket (MB), indicating the tasks essential for AD detection. By employing the MB to select the most informative subset of tasks, our approach provides a more reliable prediction of AD onset. This study not only underscores the importance of multimodal fusion in replicating the complex decision-making process of clinical experts but also demonstrates improved performance over single-task and unimodal classifiers.

II. DATA AND ACQUISITION PROTOCOL

The protocol consists of 25 handwriting tasks arranged in increasing order of difficulty based on the cognitive functions required to complete them. Considering their specific objectives, we have categorised the tasks as follows:

- **Graphic tasks**: six tasks designed to assess the patient's ability in writing basic strokes, connecting points, and drawing/retracing figures, which can range from simple to complex and are scaled in various sizes;
- **Copy and Reverse Copy tasks**: fourteen tasks designed to assess the patient's ability to replicate complex graphic gestures with semantic meaning, such as letters, words, and numbers, varying in length and spatial arrangement;

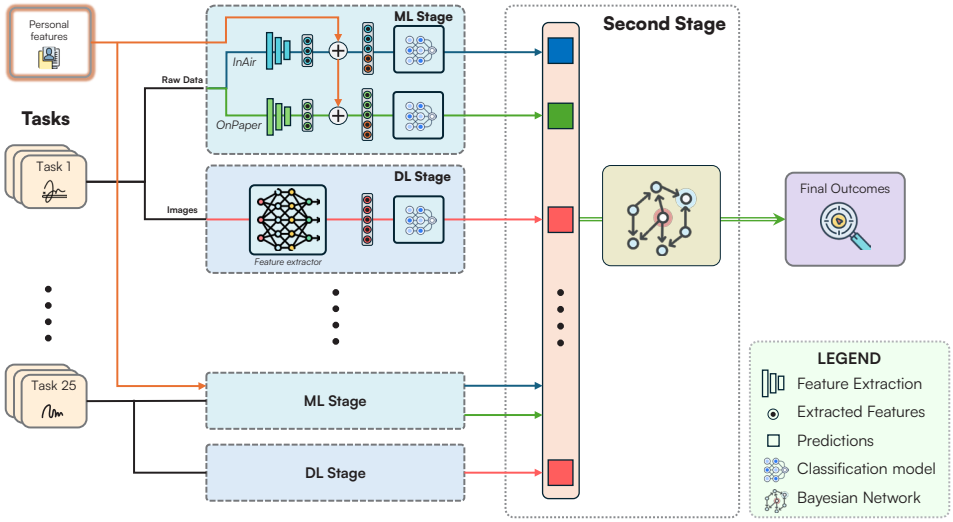


Fig. 1. Scheme of the proposed approach using early fusion technique.

- **Memory task**: one task focused on testing how the patient's handwriting changes while recalling words;
- **Dictation task**: four tasks that examine how writing phrases or numbers are affected when working memory is required.

The protocol was executed by 174 participants, including 89 AD patients (AP) and 85 healthy controls (HC), framing our analysis as a binary classification problem. The two groups are matched for age, gender, and education. The AD participants were patients from the Alzheimer's Disease Unit at the Federico II hospital in Naples. AD and HC were selected based on their preserved cognitive abilities, necessary to complete the protocol and restricted recruitment criteria as detailed in [12]. Thanks to the Wacom Bamboo Folio, we acquired handwriting samples, allowing people to write with a sensor-equipped pen on a sheet of paper fastened to the tablet surface. The tool acquired (x, y) spatial coordinates, the pressure (z) and the timestamp at a sampling frequency of $200Hz$. In addition, air movements were detected within 3 cm of the tablet surface. We extract static and dynamic features from the handwriting data for each subject and task. Static features, like size and slant, are time-invariant, while dynamic features, such as velocity and pressure change, capture temporal characteristics. As many studies highlight significant differences in patients' motor performance between in-air and on-paper traits, each feature was extracted separately for in-air and on-paper traits. In particular, the **OnPaper** features were extracted from the written traits, while the **InAir** features were extracted from the in-air traits only. These latter movements characterise the planning activity for positioning the pen tip between two successive written traits; More details on these two groups of features and the other two groups (All and In-air-on-paper), not used in this paper, can be found in [13]. Finally, we complete these datasets with the subject's personal features (age, gender, and education, intended as school years). Such additional features also affect performance, adding a marginal complexity to the model. In addition to these features, we also used offline handwriting images acquired for each task and applied CNNs to extract features from the pixel values of the natural ink shades on the paper. This approach effectively captures variations in handwriting shape, size, and thickness. Even though we lose the movement dynamics, the image allows us to better assess the patient's spatial awareness.

III. The Proposed Approach

The proposed architecture, depicted in Figure 1, outlines the workflow for processing data from a single subject. It consists of two stages: the first stage focuses on classification using various feature spaces, while the second stage combines classifier predictions. In the first stage, each task is processed through three distinct channels. Two channels use the online handwriting raw data described in Section II during the ML Stage, one for **InAir** features and one for **OnPaper** features. Each feature group is fused with the subject's personal information to produce a unique feature vector. The third channel uses the offline handwriting image of each task and applies a CNN to extract features. At this point, a classification model is trained for each feature space, yielding a prediction and an associated confidence score for each channel. Since this is done for each of the 25 tasks across three channels, the process generates a prediction vector containing 75 entries (25 tasks × 3 channels).

In the second stage, these 75 predictions and the class label are used to train a BN. As mentioned in the introduction, a BN is a probabilistic graphical model representing the joint probability of random variables using a DAG, where nodes denote the variables and arrows indicate their statistical dependencies. In our case, with the BN learning we aim to estimate the joint probability of random variables associated with the class label (l) and the 75 tasks (t_{IA1}, t_{OP1}, t_{CNN1}, …, t_{IA25}, t_{OP25}, t_{CNN25}). In the DAG, nodes represent these variables with arrows indicating conditional dependencies, where t_j can be defined

as a *parent* of t_i (i.e., $t_j \in pa_{t_i}$). For each node, either a conditional probability $p(t_i|pa_{t_i})$ quantifies parental influence or an a priori probability $p(t_i)$ indicates conditional independence. The BN learning captures both network structure (statistical dependencies) and probability distribution parameters using constraint-based, score-based, or hybrid methods. After structural learning, parameter learning enables the association of the DAG with the joint probability:

$$p(l, t_1, \ldots, t_{75}) = p(l|pa_l) \prod_{t_i \in T_l} p(t_i|pa_{t_i}) \prod_{t_i \in \bar{T}_l} p(t_i|pa_{t_i}) \tag{1}$$

In Eq.1, we distinguish tasks directly linked to the class label (pa_l and $t_i \in T_l$), known as the Markov Blanket (MB), from other tasks $t_i \in \bar{T}_i$ that neither have the class label as parent nor are parents of the class label. Thus, in Eq.1, we can distinguish tasks that are directly linked to the class label (pa_l and $t_i \in T_l$ also comprising the t_i parents), also known as Markov Blanket (MB), from the other tasks $t_i \in \bar{T}_i$ that are neither *parent* of the class label neither has the class label as its parent $\bar{T}_i$. The MB allows the identification of a set of sources, in our case, the handwriting tasks, in a multimodal representation useful for analysing the handwriting data. The selection performed through structural learning allows us to consider only the tasks that interact in the AD detection process simultaneously and in a complementary way. Thus, the trained BN provides: (i) a second-level classification of AD onset, exploiting direct inference, and (ii) a task selection encoded in the MB. Additionally, the confidence degrees associated with each task classifier are used as weights for implementing the combining methods.

IV. Experimental study

We implemented our approach using the scikit-learn library [14]. For the CNN-based feature channel, we employed the ConvNeXt Small architecture [15], which is a CNN inspired by transformer architectures. It consists of 50 million parameters with hierarchical stages of convolutions and LayerNorm layers, using 3×3 depthwise convolutions and GELU activations. For each task and feature channel (OnPaper, InAir, CNN), we follow the process:

1) We performed 5-fold stratified cross-validation on the training set to ensure robust evaluation and improve model generalisation. To prevent data leakage and maintain consistency, we ensured that subjects were equally distributed across all tasks for each run. This meant that a given subject's data was consistently assigned to the training, validation, or test set across all tasks.
2) We selected Decision Tree (DT) [16], Random Forest (RF) [17], Support Vector Classifier (SVC) [18], or XGBoost (XGB) [19] as base classifiers due to their diverse decision-making processes and complementary strengths. In each run, all base classifiers were the same type.
3) We conducted hyperparameter tuning for each classifier using the Optuna optimisation framework [20], specifically employing Optuna's Tree-structured Parzen Estimator (TPE) Sampler and using 100 trials.
4) After hyperparameter tuning, we trained the models and obtained predictions with confidence scores. We then calibrated these confidence scores using isotonic regression to correct overly confident or underconfident predictions, improving the model's decision-making reliability [21].
5) We merged the tasks' predictions with the class label to train a BN following the supervised learning procedures described in Section III. We implemented a BN using *pgmpy* Python library [22], with BDeu score and Hill Climbing structural learning algorithm. We set the hyperparameter $\alpha = 1$, and the maximum number of parent constraints equal to seven.
6) Finally, we carried out three types of results on the test set: (i) direct inference from the BN, (ii) Majority Vote using tasks selected by the MB (MV_MB), and (iii) Weighted Majority Vote using tasks selected by the MB, where weights were based on the confidence scores obtained in the previous classification step (WMV_MB).

We repeated this entire process for 30 independent runs to ensure statistical robustness. This approach allowed us to assess the consistency and reliability of our results across multiple iterations. We used three metrics to evaluate the proposed system's performance: accuracy, sensitivity, and specificity. The accuracy (Acc) measures the overall effectiveness, whereas sensitivity (Sen) focuses on the model's ability to correctly identify AD patients, and specificity (Spe) evaluates the model's capability to classify healthy individuals accurately. In medical diagnostics, sensitivity is vital for accurately identifying AD patients. Table I shows the mean and standard deviation of those metrics calculated over 30 runs. The four triple columns depict the results obtained from the prediction datasets given by the base classifiers tested (DT, RF, SVC, XGB). Each row reports the results of BN, MV_MB, and WMV_MB tested methods compared with two classical ensemble techniques (MV and WMV) and two stacking methods that employed Multilayer Perceptron and Logistic Regression as meta-learners (Stck_MLP and Stck_LR, respectively). In the last two rows, we present the single best (SB) classifier result and indicate the task and channel that achieved the highest accuracy.

These preliminary experiments indicate that the highest overall result of **89.56%** was achieved using WMV_MB combining method with the XGB classifiers, showing notable improvement over our previous work [7], which achieved 84.21% using only In-Air-On-Paper features with XGB. This improvement demonstrates the benefit of integrating CNN-processed offline images, though we observed that the CNN channel occasionally introduces imbalances between sensitivity and specificity

TABLE I. MEAN AND STANDARD DEVIATION OF ACCURACY, SENSITIVITY, AND SPECIFICITY ON EARLY FUSED DATA USING DIFFERENT SECOND-STAGE METHODS.

	SVC			**DT**			**RF**			**XGB**		
	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
BN	72.08 (6.37)	65.48 (6.92)	98.13 (5.22)	86.60 (4.13)	89.05 (3.78)	89.20 (5.27)	76.04 (3.12)	72.26 (5.17)	91.87 (2.03)	85.47 (1.17)	82.38 (0.87)	93.33 (1.53)
MV_BN	86.92 (5.73)	76.90 (6.43)	79.47 (6.83)	89.18 (3.23)	89.17 (4.32)	83.87 (5.63)	81.38 (2.57)	72.02 (4.87)	80.27 (3.83)	**89.56 (0.97)**	86.19 (1.23)	88.93 (1.03)
WMV_BN	55.47 (7.02)	99.17 (6.88)	06.53 (7.04)	78.24 (4.57)	95.36 (2.63)	59.07 (4.17)	67.42 (5.62)	97.74 (4.12)	33.47 (5.23)	75.35 (1.83)	95.71 (0.73)	52.53 (1.47)
MV	84.65 (5.23)	77.26 (6.32)	92.93 (5.67)	76.48 (5.03)	74.29 (4.83)	78.93 (3.73)	76.35 (3.67)	65.83 (5.23)	88.13 (2.33)	77.36 (1.63)	67.50 (1.33)	88.40 (1.73)
WMV	86.35 (4.83)	94.17 (6.93)	77.60 (5.33)	79.50 (4.13)	83.81 (4.23)	74.67 (5.33)	76.16 (4.22)	88.10 (3.03)	62.80 (5.13)	80.75 (1.03)	83.21 (0.63)	78.00 (1.23)
Stck_LR	68.05 (7.04)	69.86 (6.43)	61.42 (6.77)	81.00 (3.13)	86.74 (2.53)	71.83 (4.23)	77.24 (3.92)	76.81 (4.43)	76.42 (4.83)	82.95 (0.83)	82.06 (1.13)	84.08 (0.93)
Stck_MLP	59.81 (7.06)	58.44 (7.02)	65.92 (6.12)	81.07 (3.43)	88.94 (2.03)	74.25 (4.53)	76.63 (4.23)	76.81 (5.13)	77.75 (3.73)	83.45 (1.33)	82.91 (0.53)	84.00 (1.63)
Single Best	64.68 (2.31)	60.24 (3.12)	68.17 (2.57)	83.06 (1.48)	80.07 (2.44)	85.92 (1.89)	79.61 (2.05)	76.83 (2.90)	81.56 (2.40)	87.93 (1.63)	85.21 (2.01)	88.44 (1.54)

metrics, particularly with SVC models. The integration of personal features (age, education, gender) contributed approximately 3-5% improvement in accuracy across models, with education level showing the strongest correlation with task performance. The WMV_MB outperforms all other combination approaches when paired with both high-performing classifiers like XGB (which has an individual performance of 87.93%) and lower-performing ones such as SVC (64.68%). In two out of four cases, WMV_MB demonstrated a 10-percentage point improvement over the classical WMV approach, highlighting the critical role of task selection via the Markov Blanket. Furthermore, the BN approach outperforms stacking methods, which typically perform well only when built upon an already strong ensemble. The BN effectively exploits classifier diversity, increasing overall classification performance.

REFERENCES

[1] M. A. El-Yacoubi, S. Garcia-Salicetti, C. Kahindo, A.-S. Rigaud, and V. Cristancho-Lacroix, "From aging to early-stage alzheimer's: Uncovering handwriting multimodal behaviors by semi-supervised learning and sequential representation learning," *Pattern Recognition*, vol. 86, pp. 112–133, 2019.

[2] R. Saha, A. Mukherjee, A. Sadhukhan, A. Roy, and M. De, *Handwriting Analysis for Early Detection of Alzheimer's Disease*. Wiley Data and Cybersecurity, 2020, pp. 369–385.

[3] H. Qi, R. Zhang, Z. Wei, C. Zhang, L. Wang, Q. Lang, K. Zhang, and X. Tian, "A study of auxiliary screening for alzheimer's disease based on handwriting characteristics," *Frontiers in aging neuroscience*, vol. 13, p. 1117250, 2023.

[4] N. D. Cilia, C. De Stefano, F. Fontanella, and A. Scotto di Freca, "An experimental protocol to support cognitive impairment diagnosis by using handwriting analysis," *Procedia Computer Science*, vol. 141, pp. 466 – 471, 2018.

[5] N. Cilia, C. De Stefano, F. Fontanella, and A. Scotto di Freca, "Feature selection as a tool to support the diagnosis of cognitive impairments through handwriting analysis," *IEEE Access*, vol. 9, pp. 78 226–78 240, 2021.

[6] N. Dalia Cilia, T. D'Alessandro, C. De Stefano, and F. Fontanella, "Offline handwriting image analysis to predict alzheimer's disease via deep learning," in *2022 26th International Conference on Pattern Recognition (ICPR)*, 2022, pp. 2807–2813.

[7] E. Nardone, T. D'Alessandro, C. De Stefano, F. Fontanella, and A. Scotto di Freca, "A bayesian network combiner for multimodal handwriting analysis in alzheimer's disease detection," *Pattern Recognition Letters*, vol. 190, pp. 177–184, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865525000613

[8] A. Kline, H. Wang, Y. Li, S. Dennis, M. Hutch, Z. Xu, F. Wang, F. Cheng, and Y. Luo, "Multimodal machine learning in precision health: A scoping review," *npj Digital Medicine*, vol. 5, no. 1, p. 171, Nov 2022.

[9] E. Kyrimi, S. McLachlan, K. Dube, M. R. Neves, A. Fahmi, and N. Fenton, "A comprehensive scoping review of bayesian networks in healthcare: Past, present and future," *Artificial Intelligence in Medicine*, vol. 117, p. 102108, 2021.

[10] D. R. Faria, C. Premebida, L. J. Manso, E. P. Ribeiro, and P. Núñez, Multimodal Bayesian Network for Artificial Perception," *Bayesian Networks - Advances and Novel Applications*, D. McNair, Ed. Rijeka: IntechOpen, 2018, ch. 6.

[11] H. Irzan, M. Hütel, H. O'Reilly, S. Ourselin, N. Marlow, and A. Melbourne, "Multi-source multi-modal markers for bayesian networks: Application to the extremely preterm born brain," *Medical Image Analysis*, vol. 92, p. 103037, 2024.

[12] N. D. Cilia, T. D'Alessandro, C. De Stefano, F. Fontanella, and M. Molinara, "From online handwriting to synthetic images for alzheimer's disease detection using a deep transfer learning approach," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 12, pp. 4243–4254, 2021.

[13] T. D'Alessandro, C. De Stefano, F. Fontanella, E. Nardone, and A. Scotto di Freca, "Feature evaluation in handwriting analysis for alzheimer's disease using bayesian network," in *Graphonomics in Human Body Movement.*, A. Parziale, M. Diaz, and F. Melo, Eds. Springer Nature Switzerland, 2023, pp. 122–135.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[15] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," 2022. [Online]. Available: https://arxiv.org/abs/2201.03545

[16] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.

[17] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[18] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[19] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 2016, pp. 785–794.

[20] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[21] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proceedings of the 34th International Conference on Machine Learning*. PMLR, 2017, pp. 1321–1330.

[22] A. Ankan and A. Panda, "pgmpy: Probabilistic graphical models using python," in *Proceedings of the Python in Science Conference*, ser. SciPy. SciPy, 2015.





Session

Student Competition

Biceps Curl Exercise Analysis with Wearable Smartwatch Sensors

Cristian Rodriguez, David Espino, Cristina Jimenez-Marcos, Moises Diaz, Cristina Carmona-Duarte, Miguel A. Ferrer

iDeTIC, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain

{cristian.rodriguez110, cristina.jimenez112}@alu.ulpgc.es, {david.espino, moises.diaz, cristina.carmona, miguelangel.ferrer}@ulpgc.es

***Abstract*— Strength training, especially the biceps curl, is key to improving arm and elbow function by enhancing muscle strength and mobility. But how can we accurately determine if each repetition is performed correctly? This study tackles that question by monitoring angular velocity and the duration of concentric and eccentric phases using the inertial measurement unit (IMU) of a smartwatch. We worked with 100 participants of varying physical conditions, health statuses, and exercise backgrounds. Their biceps curl repetitions were recorded and stored in a labeled database. Using this data, we trained three machine learning models to predict whether each repetition was executed properly. The results are promising and point to the potential for developing an app that could support coaches, beginners, and experienced athletes in optimizing their training, improving efficiency, and reducing the risk of injury.**

Keywords*— *Sport applications, Angular velocity, Strength training, Biceps Curl, Machine Learning

I. Introduction

Strength exercises are essential for building muscle, increasing strength, and preventing mobility and postural issues over time. One of the most effective exercises in this area is the subject of this study, the biceps curl. This exercise not only strengthens the arm muscles but also enhances elbow mobility and flexibility. However, doing it incorrectly can reduce its benefits and even cause mobility problems in the arm and elbow.

While most research on the biceps curl has focused on high-performance athletes, limited attention has been given to factors that can influence exercise technique, such as the individual's experience level, gender differences, dominant hand preference, and overall physical health. To address this gap, this study examines these variables by analyzing biomechanical parameters like execution time, angular velocity, and motor control during the exercise. The goal is to develop machine learning models capable of accurately monitoring whether the exercise is performed correctly, regardless of the individual's characteristics. The biceps curl consists of two phases: the concentric phase, where the dumbbell is lifted, and the eccentric phase, where it is lowered (see Fig.1, left). Proper understanding of both phases is essential for assessing form and optimizing training [1].

Studies have shown that the time spent in the eccentric phase significantly impacts muscle development [2]. For example, durations longer than 3 seconds in this phase can increase muscle mass by 15–20% compared to shorter durations. Further research recommends a total repetition time of 2 to 10 seconds for optimal muscle growth, with eccentric phases lasting 3–5 seconds leading to greater muscle activation [3]. Studies have also demonstrated that the timing of eccentric phases is crucial for improving muscle strength and development [4]. Despite substantial research on the biceps curl, there is still a lack of real-time devices to assess repetition quality. Variations in physical conditions, like experience or age, make it challenging to apply a universal approach. More studies are needed to determine the best techniques for different profiles to optimize muscle development and minimize injury risk [5].

This study is divided into five sections: the introduction, which reviews the motivation and relevant literature; the description of the database, including participant demographics and experimental parameters; the methodology, which explains the development of a classifier using machine learning; the results, which showcase model performance; and the conclusion, which analyzes the outcomes and summarizes key findings.

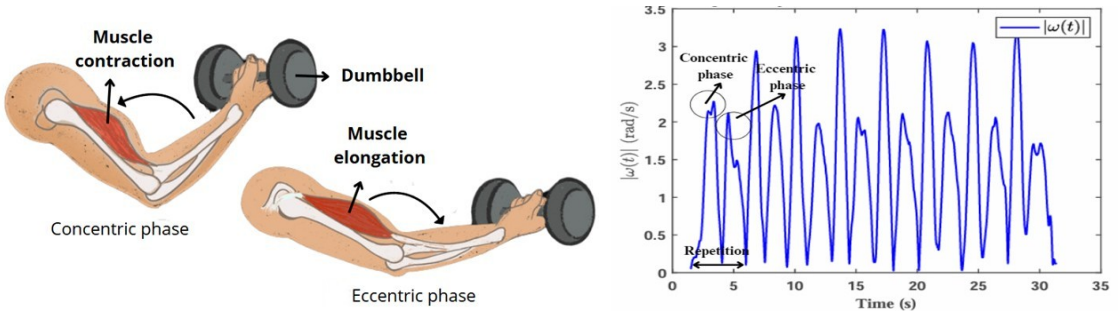


Fig. 1: Execution of the concentric and eccentric phase (left), and angular velocity variation during exercise (right).

II. Database

Data was collected from 100 volunteers aged 18 to 59, mostly between 18 and 25 years, with varied characteristics such as gender, health status, hand dominance, and training experience. Each wore two Apple Watch Series 6 devices with IMUs (accelerometer, gyroscope, and magnetometer) and performed simultaneous biceps curls with both arms using 3 kg or 5 kg dumbbells until fatigue. The watches recorded tri-axial angular velocity data ($\omega_x(t)$, $\omega_y(t)$, $\omega_z(t)$) at 100 Hz. Repetitions were labeled as '0' (incorrect) or '1' (correct), and participant data was compiled into a public database[1]. Participants who did not comply with the protocol, such as by using external support or leaning against a wall while lifting the dumbbell, were excluded to preserve the methodological integrity of the study. Table I summarizes participant demographics.

TABLE I: Statistical distribution of the participants

Statistics	**Total**		**Men**			**Women**		
data	***N°***[1]	***Rep***[4]	***N°***[1]	***Rep***[4]	**%**	***N°***[1]	***Rep***[4]	**%**
Total Participants	100	6246	68	4710	68	32	1536	32
Right-Handed	91	5542	62	4114	68.1	29	1428	31.9
Left-Handed	9	704	6	596	66.7	3	108	33.3
Experienced Arm Exercise[2]	56	3916	41	3190	73.2	15	726	26.8
Non-Experienced Arm Exercise[3]	44	2330	27	1520	61.4	17	810	38.6
Injuries by Right Arm	10	632	5	326	50	5	306	50
Injuries by Left Arm	13	765	8	541	61.5	5	224	38.5

[1]Number of participants. [2]Person with experience in strength training.
[3]Person with no experience in strength training. [4]Number of repetitions.

To determine whether a biceps curl was performed correctly, sensor data from each arm was used to calculate angular velocity as follows, $|\omega(t)| = \sqrt{\omega_x^2(t) + w_y^2(t) + w_z^2(t)}$. The expected signal follows a sinusoidal pattern, where each period represents a phase of the exercise, with two successive peaks indicating one complete repetition (see Fig.1, right).

From $|\omega(t)|$, each participant's repetitions were segmented by detecting the zero-crossing points of $\omega_x(t)$ and applying a low-pass filter to reduce noise. The concentric and eccentric phases were then separated, as illustrated in Fig. 3. To evaluate whether a repetition was performed correctly, several angular velocity parameters were extracted for each repetition, and a classifier was applied to assess the results.

A. Parameters

Several parameters are extracted from the angular velocity modulus to assess whether a biceps curl repetition is performed correctly, with separate analysis for the concentric and eccentric phases.

1) Number of peaks: The number of peaks in the concentric ($|\omega_{cc}(t)|$) and eccentric ($|\omega_{ecc}(t)|$) phases indicates exercise control. More than one peak suggests poor control, while a single peak signals proper technique, as shown in Fig.2 [5], [6].

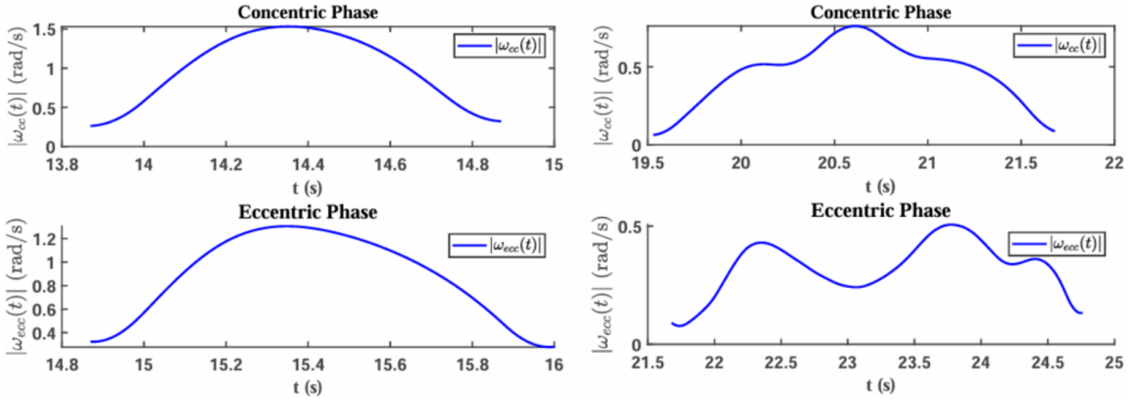


Fig. 2: Segmented trajectory of the $|\omega(t)|$ in the concentric and eccentric phases

2) Absolute maximum peak value: The concentric phase is characterized by an explosive movement requiring significant muscular strength and speed, whereas the eccentric phase focuses on control and resistance, involving slower movement. Consequently, the angular velocity modulus in the concentric phase should be at least 10% higher than in the eccentric phase, i.e. $max|\omega_{cc}(t)| > 1.1 \cdot max|\omega_{ecc}(t)|$ [6].

[1]https://gpds.ulpgc.es [Accessed: Apr. 29, 2025]

3) Time of the concentric and eccentric phase: The concentric phase, being explosive, must have a shorter duration, while the eccentric phase, which requires greater control, should last longer. Thus, the eccentric phase duration should be at least 10% longer than the concentric phase, i.e. $t_{ecc} > 1.1 \cdot t_{cc}$. The 10% threshold between angular velocity and time was empirically defined as a tolerance margin to distinguish controlled eccentric phases from explosive concentric ones.

4) Total Repetition Time: The total repetition time, $t_{rep} = t_{cc} + t_{ecc}$, for a repetition to be considered correct, should range between 2 and 8 seconds. This ensures efficiency, prevent fatigue, and reduce injury risk: $2 \leq t_{rep} \leq 8$.

5) Lognormal fit: The lognormal fit is widely used to analyze the velocity of rapid human movements. In this study, it is compared with the angular velocity modulus curve. The lognormal function is defined as 1.

$$f_{log}(t) = \frac{1}{|\omega(t)|\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(|\omega(t)| - \mu)^2}{2\sigma}\right) \tag{1}$$

where μ represents the mean of the logarithm of $|\omega(t)|$ and σ is the standard deviation of the natural logarithm of $|\omega(t)|$. Once the lognormal curve and the angular velocity modulus curve are fitted, the area difference between the two curves is analyzed to evaluate exercise efficiency in both the concentric ($\Delta Area_{cc}$) and eccentric ($\Delta Area_{ecc}$) phases, as defined by $\Delta Area_{phase} = \int_0^{t_{phase}} |f_{log}(t) - f_{|w|}(t)|dt$.

For correct execution, the concentric phase should align closely with the lognormal curve, while the eccentric phase should deviate, indicating controlled movement [2], [4]. A graphical representation of the area difference is shown in Fig.3, with the deviation highlighted in green. The lognormal probability density function "lognpdf" is used to fit $|\omega(t)|$ with $f_{log}(t)$, optimizing μ and σ to minimize the quadratic difference using the "fminsearch" function. The area under the curve is then computed to compare the deviation between the actual data and the fitted distribution.

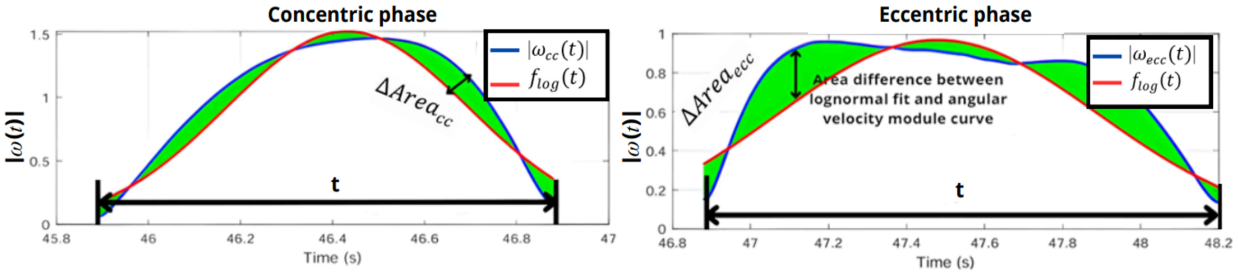


Fig. 3: Area difference between lognormal fit and angular velocity modulus curve in concentric and eccentric phases

III. Bicep Curl Evaluation

This study evaluates biceps curl repetitions using machine learning techniques: Support Vector Machine (SVM), Neural Network (NN), and Random Forest (RF). SVM identifies a hyperplane that maximizes class separation, NN detects patterns with interconnected nodes, and RF creates decision trees to improve accuracy through majority voting.

The classifiers were trained on 6,246 repetitions from 100 participants, with input features including angular velocity $|\omega_{phase}(t)|$, number of peaks $max(|\omega_{phase}(t)|)$, execution time (t_{phase}) and lognormal area ($\Delta Area_{phase}$). The data were split into training (90%) and testing (10%) sets, and the models were trained using MATLAB's "fitcsvm," "fitnet," and "TreeBagger" functions. No specific hyperparameter optimization was performed. Future work may explore tuning these parameters to further improve classification accuracy.

Model performance was evaluated based on accuracy (ACC), which measures the proportion of correctly classified repetitions compared to visual classifications; area Under the Curve (AUC), which assesses how effectively the model distinguishes between correct and incorrect repetitions; sensitivity, Se, representing the proportion of correctly executed repetitions identified by the model and calculated as: $Se = 2 \cdot \frac{TP}{TP+FN}$, where TP and FN refer to True Positive and False Negative, respectively. F1-score, $F1$, calculated as: $F1 = \frac{ACC \cdot Se}{ACC+Se}$, providing a weighted balance of accuracy and sensitivity, particularly useful in cases of class imbalance; and specificity Sp, , that indicates the proportion of poorly executed repetitions correctly identified by the model and is defined as: $Sp = \frac{TN}{TN+FP}$, where TN and FP denote True Negative and False Positive, respectively [7].

IV. Results

A comparison was made between the classifiers (SVM, NN, and RF) using varying training data proportions (10% to 90%) to assess classification accuracy. NN and RF consistently outperformed SVM, achieving accuracy above 0.9, while SVM stayed below 0.85. As training data increased, all models improved, with NN and RF converging at similar high accuracy, though NN slightly exceeded RF in some cases. Stability analysis showed that NN and RF were more reliable than SVM. The results, presented in Table II, cover all subgroups with their metrics. These results correspond to training with 50% of the data as the classification improved slightly with increasing data and to avoid overfitting.

TABLE II: Performance of the SVM, NN and RF for different groups

Subgroups	Method	ACC	AUC	Sensitivity	F1-score	Specificty
General (100% participants)	SVM	0.83 ± 0.01	0.91 ± 0.01	0.58 ± 0.02	0.65 ± 0.01	0.92 ± 0.01
	NN	0.95 ± 0.01	0.99 ± 0.01	0.92 ± 0.01	0.91 ± 0.01	0.96 ± 0.01
	RF	0.94 ± 0.01	0.89 ± 0.01	0.86 ± 0.01	0.88 ± 0.01	0.97 ± 0.01
Male (68% participants)	SVM	0.85 ± 0.01	0.92 ± 0.01	0.64 ± 0.02	0.69 ± 0.02	0.93 ± 0.01
	NN	0.95 ± 0.01	0.98 ± 0.01	0.90 ± 0.02	0.90 ± 0.02	0.96 ± 0.01
	RF	0.94 ± 0.01	0.98 ± 0.01	0.87 ± 0.01	0.88 ± 0.02	0.97 ± 0.01
Female (32% participants)	SVM	0.83 ± 0.01	0.91 ± 0.01	0.58 ± 0.01	0.62 ± 0.02	0.91 ± 0.01
	NN	0.93 ± 0.01	0.97 ± 0.01	0.87 ± 0.07	0.86 ± 0.04	0.95 ± 0.02
	RF	0.92 ± 0.01	0.96 ± 0.01	0.76 ± 0.04	0.82 ± 0.02	0.97 ± 0.01
Right-handed (91% participants)	SVM	0.83 ± 0.01	0.91 ± 0.01	0.62 ± 0.02	0.67 ± 0.01	0.91 ± 0.01
	NN	0.94 ± 0.01	0.98 ± 0.01	0.93 ± 0.02	0.91 ± 0.01	0.96 ± 0.01
	RF	0.94 ± 0.01	0.98 ± 0.01	0.87 ± 0.01	0.89 ± 0.01	0.96 ± 0.01
Left-handed (9% participants)	SVM	0.92 ± 0.01	0.94 ± 0.02	0.72 ± 0.03	0.75 ± 0.04	0.93 ± 0.01
	NN	0.94 ± 0.01	0.95 ± 0.03	0.60 ± 0.18	0.63 ± 0.15	0.98 ± 0.01
	RF	0.95 ± 0.01	0.96 ± 0.01	0.58 ± 0.06	0.66 ± 0.06	0.98 ± 0.01
Experienced Arm Exercise (56% participants)	SVM	0.85 ± 0.01	0.96 ± 0.01	0.61 ± 0.03	0.66 ± 0.02	0.93 ± 0.01
	NN	0.94 ± 0.01	0.98 ± 0.01	0.89 ± 0.01	0.88 ± 0.03	0.96 ± 0.01
	RF	0.94 ± 0.01	0.98 ± 0.01	0.86 ± 0.02	0.88 ± 0.01	0.97 ± 0.01
Novice Arm Exercise (44% participants)	SVM	0.86 ± 0.01	0.93 ± 0.01	0.77 ± 0.03	0.77 ± 0.02	0.90 ± 0.01
	NN	0.93 ± 0.01	0.98 ± 0.01	0.91 ± 0.03	0.89 ± 0.02	0.94 ± 0.01
	RF	0.93 ± 0.01	0.97 ± 0.01	0.84 ± 0.02	0.87 ± 0.01	0.95 ± 0.01
Right Arm Injury (10% participants)	SVM	0.85 ± 0.01	0.92 ± 0.01	0.69 ± 0.08	0.71 ± 0.04	0.91 ± 0.02
	NN	0.89 ± 0.02	0.95 ± 0.02	0.80 ± 0.07	0.81 ± 0.04	0.94 ± 0.02
	RF	0.90 ± 0.01	0.95 ± 0.01	0.74 ± 0.05	0.80 ± 0.03	0.96 ± 0.01
Left Arm Injury (13% participants)	SVM	0.86 ± 0.01	0.94 ± 0.01	0.64 ± 0.06	0.70 ± 0.04	0.93 ± 0.01
	NN	0.91 ± 0.01	0.96 ± 0.02	0.81 ± 0.17	0.80 ± 0.14	0.94 ± 0.02
	RF	0.91 ± 0.01	0.97 ± 0.01	0.81 ± 0.06	0.82 ± 0.03	0.94 ± 0.01

V. Conclusion

When analyzing the performance of the biceps curl exercise, the results show that both Neural Networks (NN) and Random Forests (RF) achieve acceptable performance, exceeding 90% in most cases. These findings suggest that these classifiers are effective for evaluating and scoring physical exercises like the Biceps Curl in an automated and transparent way. However, lower performance was observed in participants with a history of injury, an area that requires further investigation in future studies. NN and RF demonstrate strong potential in sports and clinical settings, where accurate assessment of exercise technique is crucial. Conversely, Support Vector Machines (SVM) showed one of the lowest performances, indicating it may not be the best fit for this study's data. For future research, we aim to explore additional parameters and end-to-end learning approaches using angular velocity time series and other derived physical parameters like the dynamics through the calculus of force with Lagrangian formulation and robotic arm models. Our ultimate goal is to develop a smartwatch app that provides real-time feedback to optimize performance and reduce injury risk, this is useful not only for athletes, but also for older adults who want to maintain mobility and strength.

Acknowledgment

This research was supported by the PID2023-146620OB-I00, funded by MICIU/AEI 10.13039/501100011033, and the European Union's FEDER program and the CajaCanaria and la Caixa (2023DIG05).

References

[1] M. Ansari, S. Hardcastle, S. Myers, and A. D. Williams, "The health and functional benefits of eccentric versus concentric exercise training: A systematic review and meta-analysis," *JOURNAL OF SPORTS SCIENCE AND MEDICINE*, vol. 22, pp. 288–309, JUN 2023.

[2] F. Kojić, I. Ranisavljev, D. Ćosić, D. Popović, S. Stojiljković, and V. Ilić, "Effects of resistance training on hypertrophy, strength and tensiomyography parameters of elbow flexors: role of eccentric phase duration," *Biology of Sport*, vol. 38, no. 4, pp. 587–594, 2021.

[3] B. J. Schoenfeld, D. I. Ogborn, and J. W. Krieger, "Effect of repetition duration during resistance training on muscle hypertrophy: a systematic review and meta-analysis," *Sports Medicine*, vol. 45, pp. 577–585, 2015.

[4] A. Laurent, R. Plamondon, and M. Begon, "Central and peripheral shoulder fatigue pre-screening using the sigma–lognormal model: A proof of concept," *Frontiers in human neuroscience*, vol. 14, p. 171, 2020.

[5] S. Boyas, A. Remaud, E. Rivers, and M. Bilodeau, "Fatiguing exercise intensity influences the relationship between parameters reflecting neuromuscular function and postural control variables," *PloS one*, vol. 8, no. 8, p. e72482, 2013.

[6] A. C. Sy *et al.*, "Effects of movement velocity, acceleration and initial degree of muscle flexion on bicep emg signal amplitude," in *Int. Conf. on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, pp. 1–6, IEEE, 2015.

[7] S. Albahra *et al.*, "Artificial intelligence and machine learning overview in pathology & laboratory medicine: A general review of data preprocessing and basic supervised concepts," *SEMINARS IN DIAGNOSTIC PATHOLOGY*, vol. 40, pp. 71–87, MAR 2023.

The Effect of Keyboard Size and Anthropometry on Pianists' Finger Kinematics

Craig Turner[1,2], Robin Mailly[1,2], Justine Pelletier[3], Mickaël Begon[1], Felipe Verdugo[1,2]

[1]*Laboratoire de Simulation et Modélisation du Mouvement, École de Kinésiologie et des Sciences de l'Activité Physique, Université de Montréal, Québec, Canada*

[2]*Centre de Recherche Interdisciplinaire en Réadaptation, Université de Montréal, Québec, Canada*

[3]*Faculté de Musique, Université de Montréal, Québec, Canada*

craig.turner@umontreal.ca

***Abstract*—**Musculoskeletal injury rates are high in pianists. Having a small hand span is a risk factor. Ergonomically smaller keyboards (e.g., 6.0" octave keyboard; 1 octave = 8 notes) might reduce risk of injury for pianists with smaller hands. No existing research has investigated the impact of keyboard size on individual finger joint kinematics. The objective of this study was to determine the effect of keyboard and hand size on pianists' finger joint angle posture and range of motion (ROM). An optoelectronic system recorded the finger kinematics of 24 pianists while they played a short musical excerpt involving large handspan chords on a conventional 6.5" octave keyboard. Fourteen participants also played on a 6.0" octave keyboard. Finger angle postures (50th, 5th, and 95th percentiles) were assessed during specific keystroke times while ROM was assessed across the entire excerpt. Participants playing on the conventional keyboard were divided into small and large hand span groups. The smaller hand group exhibited extended (rather than flexed) 3rd finger postures, and reduced 5th finger abduction ROM compared to larger pianists, suggesting greater postural constraints. For smaller pianists playing on both keyboards, most joints exhibited more flexed, less extended, or less abducted finger postures when playing on the 6.0" keyboard, in addition to greater 5th finger abduction ROM. Playing on a 6.0" octave keyboard may therefore help pianists with smaller hands reduce their exposure to risks of injury by facilitating more neutral finger postures when performing fast passages with chords encompassing an octave, which are common in piano repertoire.

Keywords—piano performance biomechanics, musculoskeletal disorders, keyboard size, hand span, finger joint angles

I. Introduction

Performance-related musculoskeletal disorders (PRMDs) affect up to 82.6% of pianists over their lifetime [1]. PRMDs commonly occur in the forearms and wrists [1], [2]. Lateral epicondylitis and *de Quervain's* syndrome are the most common types of distal injuries [3]. One prominent risk factor is small hand size [4], [5], which affects female pianists to a greater extent [6], [7]. Hand span is reported to be negatively correlated with hand PRMDs [8] and pain [9], [10]. Pianists with a small hand span exhibit a greater angle between the thumb and 3rd finger when playing repetitive octaves (1 octave spans 8 notes = 6.5" on a conventional keyboard) [11]. They also have larger angles between adjacent fingers during octaves and chords [12]. These finger postures, when sustained over long periods of time during repetitive practice, could lead to acute and chronic injuries [3]. Large handspan chords (which are common in piano repertoire), in conjunction with playing loud and fast, are risk factors of PRMDs for pianists [13]. Since having a small hand span is a non-modifiable risk factor of PRMDs, reducing exposure to PRMDs remains a challenge in the piano field.

"Fitting the task to the person" could reduce exposure to risks of PRMDs for pianists with a small hand span. Since piano keyboards have fixed dimensions, pianists with a small hand span must adapt their movements and postures to accommodate these constraints or substantially limit their repertoire to avoid large chords and arpeggios. Ergonomically smaller keyboards have been recommended to help pianists with a small hand span [6]. However, only a few studies have assessed the impact of these keyboards on pianists' movements. In a case study on two pianists, Wristen and colleagues [14] found that playing on a 5.5" octave keyboard reduced hand span angles (the angle between the 1st phalanx of the thumb and 5th finger). In terms of muscle activity, Wristen & Hallbeck [15] reported no differences in finger and wrist flexor/extensor muscle activity for 24 pianists when comparing the 5.5" and conventional keyboards, while Chi and colleagues [16] found reduced mean forearm muscle activity in 24 pianists when comparing the 5.5" and conventional keyboards. Interestingly, Chi and colleagues found no clear differences in muscle activity between the 6.0" and conventional keyboards even though the 6.0" keyboard received higher rates of preference by participants when compared to the conventional keyboard. As the primary goal of ergonomically smaller keyboards is to minimize forced high-risk finger/hand postures in pianists with a smaller hand span, it is surprising that no research has assessed individual finger joint kinematics.

The objective of this study was to determine the effect of 1) keyboard size (conventional vs 6.0" octave keyboard in the case of pianists with a smaller hand span) and 2) hand anthropometry (smaller vs larger hands when playing on a conventional keyboard) on pianists' finger joint angle posture and range of motion (ROM) when playing loud and fast, large handspan chords in a musical excerpt from the repertoire. We hypothesized that, compared to large hand pianists, small hand pianists would exhibit more deviated (non-neutral) finger angles and less finger range of motion (ROM) in the transverse plane (parallel to the keyboard), and that playing on a 6.0" octave keyboard would result in more neutral angles and greater ROM.

II. Methods

A. Participants and Instrumentation

Fourteen pianists with a small hand span (<21.6 cm, average of a pianist population [7]) were recruited to evaluate the keyboard size effect on finger joint kinematics (11♀ & 3♂; 27.4 ± 9.9 years old; 1 left-handed). Ten additional pianists (10♂; 29.8 ± 8.3 years old) with a large hand span (>21.6 cm) were recruited to evaluate the anthropometry effect, resulting in a total of twenty-four participants for the study. All participants had expertise equivalent to at least a Bachelor of Music degree in piano performance. The research protocol was approved by the Université de Montréal ethics committee (No. 2024–5505). After being informed of the protocol and data collection procedures, participants gave written informed consent.

Ten reflective markers were placed on the following anatomical landmarks of each participant's right hand: the 1st carpometacarpal (CMC) joint, base of the 2nd and 5th metacarpals, the 1st, 2nd, 3rd, and 5th metacarpophalangeal (MCP) heads, as well as the 2nd, 3rd, and 5th proximal interphalangeal joints. Reflective markers were placed on the F#5 and B5 keys to identify keystroke times. Marker trajectories were recorded by a 6-camera VICON motion capture system (Oxford Metrics Ltd., Oxford, United Kingdom) and VICON Nexus (version 2.6) at a sampling rate of 200 Hz.

B. Experimental Protocol and Data Processing

All 24 participants played on a grand piano (Bösendorfer CEUS) with a conventional keyboard. The 14 participants with a small hand span also performed on a grand piano (Yamaha C7) equipped with a 6.0" octave keyboard (DS6.0). These participants were invited to attend a practice session (0.5-1.5 hours) to familiarize themselves with the smaller keyboard. Participants performed an excerpt from C. Debussy's *L'Isle Joyeuse* (measures 219-222) which included loud, large handspan chords at a fast performance speed of ~124 beats per minute (three known risk factors of injury), as in Yoshimura and Chesky [17]. A metronome familiarized participants with the required playing speed however, it was turned off during the performance for a more ecological playing context. The order of keyboards played was randomized, with each condition performed twice (trial A and trial B).

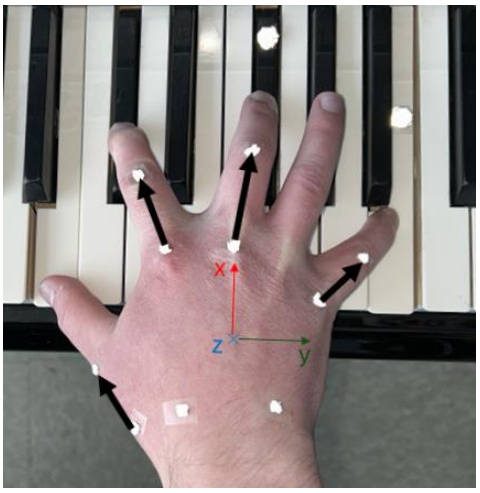


Fig. 1. Finger joint vectors and the hand's local coordinate system, with X pointing forward, Y to the right, and Z downward, while playing the chords analyzed in Debussy's *L'Isle Joyeuse* (beginning of measures 220 and 221).

Fingers were expressed in a hand local system of coordinates (Fig. 1) and converted to spherical coordinates to obtain flexion/extension and abduction/adduction joint angles for the 1st CMC, and 2nd, 3rd and 5th MCP joints. To analyze posture across all angles, two identical large handspan chord events (beginning of measures 220 and 221) were identified. These chords included an octave with additional inner notes (as shown in Fig. 1), which typically increases discomfort for pianists with smaller hands. Joint angles were analyzed 0.5 s before and 0.25 s after the selected chord events. Posture was assessed by computing the 50th (median) percentile of flexion/extension and ab/adduction angles during these chord events. The minimum (5th percentile, P5), maximum (95th percentile, P95) and ROM (P95-P5) [18] were calculated across the entire musical excerpt. The averages of each dependent variable were calculated across trials A and B.

A series of paired *t*-tests assessed the impact of keyboard size on finger posture (median, minimum, maximum angles) and ROM in the participants who played on both keyboards (N = 14; power = 0.8, d = 0.81; α = 0.05). For the anthropometry analysis, we created two distinct hand span groups among the recruited population. Small hand span was classified as <20.3 cm [15] and large hand span as >21.2 cm [17]. Six participants whose hand spans were between 20.3 cm and 21.2 cm were removed from the analysis, similar to the method used by Lai et al., [12], resulting in nine participants in the small group (hand span: 19.36 ± 0.90 cm) and nine in the large group (hand span: 22.47 ± 0.94 cm). Independent sample *t*-tests assessed the impact of anthropometry on posture variables and ROM when participants played on the conventional keyboard (N = 18; power = 0.8, d = 1.41 α = 0.05). The false discovery rate (FDR) was applied for all statistical analyses to control for potential type I errors (p = 0.05; FDR = 5%). Two participants (one from each anthropometric group) were excluded from some analysis due to a missing marker on the keyboard and hand marker occlusions.

III. Results

There was an anthropometry effect (Fig. 2) at the 3rd MCP median flexion/extension angle ($t(14) = 3.485$, $q = 0.032$), the 3rd MCP P5 flexion/extension angle ($t(14) = 3.978$, $q = 0.008$), the 5th MCP P5 flexion/extension angle ($t(14) = 3.781$, $q = 0.008$), and the 5th MCP abduction ROM ($t(14) = 3.946$, $q = 0.008$). A keyboard effect was found at the 1st CMC P5 extension angle ($t(12) = -3.415$, $q = 0.01$), the 2nd and 5th MCP's median angle and P5 ab/adduction angle (2nd MCP median: $t(12) = -4.852$, $q = 0.004$; 2nd MCP P5: $t(12) = -6.005$, $q = 0.004$; 5th MCP median: $t(12) = 3.185$, $q = 0.021$; 5th MCP P5: $t(12) = 5.025$, $q = 0.004$), the 5th MCP

This work was supported by the SSHRC Insight Development Grant [430–2021-00384], and doctoral scholarships of the 1st author: National Sciences and Engineering Research Council of Canada (NSERC) CGS D [569185–2022], and the Fonds de Recherche du Québec Nature et Technologies (FRQNT [2024–2025–B2X– 347392]).

median and P5 flexion/extension angle (median: t(12) = -4.380, q = 0.004; P5: t(12) = -3.441, q = 0.01), and the 5th MCP ab/abduction angle ROM (t(12) = -3.554, q = 0.032).

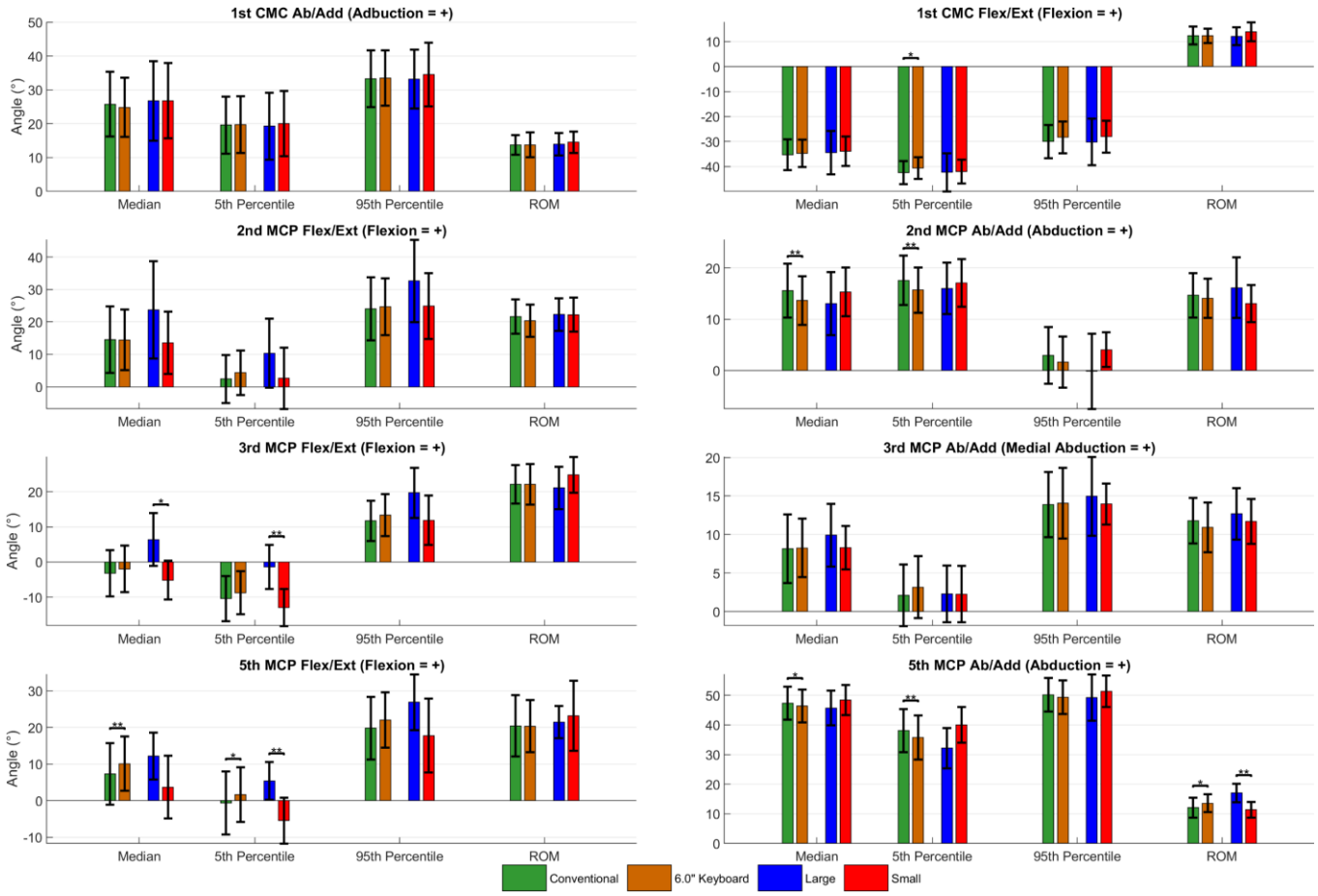


Fig. 2. Median, 5th percentile (P5), and 95th percentile (P95) angles as well as range of motion (ROM) for flexion/extension (Flex/Ext) and ab/adduction (Ab/Add) are shown for all joint angles. The keyboard effect (conventional keyboard vs 6.0" keyboard) is in green and brown while the anthropometry effect (larger vs smaller pianists) is in blue and red. Positive and negative values of the median, P5, and P95 indicate the direction of movement, with flexion as positive, extension as negative, abduction as positive, and adduction as negative. Note: * q < 0.05, ** q < 0.01.

IV. Discussion

The objective of this study was to determine the effect of keyboard and hand size on pianists' finger joint kinematics when performing loud and fast, large handspan chords encompassing an octave, which are common in piano repertoire. We found differences in 3rd and 5th MCP joint kinematics between anthropometric groups, with the small group exhibiting more extended 3rd and 5th MCP postures and reduced 5th MCP abduction ROM. 1st CMC, 2nd MCP and 5th MCP joints exhibited reduced non-neutral postures when playing on the 6.0" keyboard in addition to greater 5th finger abduction ROM.

Previous studies have shown that pianists with smaller hands, when compared to larger hands, exhibit greater thumb and 2nd finger separation [12] and increased maximum and minimum angular separation of the thumb relative to the midline of the hand [11]. In this study, no significant differences were found in 1st CMC and 2nd MCP postures between anthropometric groups, most likely because the analyzed chords constrained these fingers to remain close together. Given this chord configuration, the burden to reach all notes of the chord was placed on the opposite side of the hand. Indeed, the more extended postures detected at the 3rd and 5th MCPs, the trend of greater mean abduction postures at the 5th MCP, and the reduced 5th MCP abduction ROM in the smaller group suggest an increased stretch required from these joints, leading to their altered postures and restricted movement.

Results of the current study suggest that playing on a 6.0" octave keyboard might reduce exposure to risk factors of PRMDs for pianists with smaller hands. The 5th finger, responsible for the top note in right-hand chords, which requires greater sound intensity, showed more flexed and less abducted MCP joint angles along with greater abduction ROM on the 6.0" keyboard. These adjustments suggest that, similar to the 5.5" keyboard [14], the 6.0" keyboard can effectively help reduce extreme finger postures. The reduced 2nd MCP abduction angles suggest less stretching of the 2nd finger towards the thumb when playing these chords, a type of stretching that is usually challenging for smaller hands. Notably, although median 1st CMC posture did not differ across keyboards, the top

extension angle of this joint (represented by the 5th percentile) was reduced on the 6.0” keyboard, indicating a less extreme thumb posture towards the left. This suggests that, when playing on the 6.0” keyboard, smaller pianists may have greater opportunities to reduce thumb CMC extension angle (also called radial abduction), where even small reductions in an already extended thumb posture could decrease muscular strain. This might help reduce exposure to risk factors of PRMDs, as greater thumb extension when playing repetitive chords has been linked to *de Quervain’s* syndrome [11].

V. Limitations and Future Research

Caution is needed when interpreting the results, as a larger sample size would improve generalizability, particularly for comparisons between anthropometric groups, where only very large effect sizes could be detected. Future research could benefit from studying large chords requiring different finger configurations to enhance knowledge on the benefits of a 6.0” keyboard on finger joint kinematics in pianists with small hand spans. Including the interphalangeal joints when investigating keyboard size and anthropometry may enrich the current study’s findings. Lastly, investigating forearm muscle activity could provide insights into the relationships between the reported finger angles and muscle strain.

VI. Conclusion

Playing on a 6.0" octave keyboard may help pianists with smaller hands reduce their exposure to risks of PRMDs by allowing more neutral finger postures when performing fast passages with chords encompassing an octave, which are common in piano repertoire.

Acknowledgments

We would like to thank the International Laboratory for Brain, Music, and Sound Research (BRAMS) for providing access to the Bösendorfer CEUS piano as well as the DS Standard Foundation for providing the 6.0” octave keyboard. We also thank piano technician Marc St-Pierre for transporting and installing the 6.0” octave keyboard mechanism in our laboratory as well as Sylia Rabhi for labeling the motion capture data.

References

[1] L. Zhao, Y. Wang, and Z. Zhang, “Playing-related musculoskeletal disorders among Chinese conservatoire piano students: prevalence, risk factors and preventive interventions,” *Front. Psychol.*, vol. 15, Oct. 2024, https://doi.org/10.3389/fpsyg.2024.1386661.

[2] L. Amaral Corrêa, L. Teixeira dos Santos, E. N. Nogueira Paranhos, A. I. Minetti Albertini, P. do Carmo Silva Parreira, and L. A. Calazans Nogueira, “Prevalence and risk factors for musculoskeletal pain in keyboard musicians: A systematic review,” *PM&R*, vol. 10, no. 9, pp. 942–950, Sep. 2018, https://doi.org/10.1016/j.pmrj.2018.04.001.

[3] N. Sakai, “Hand pain attributed to overuse among professional pianists: A study of 200 cases,” *Med. Probl. Perform. Art.*, vol. 17, no. 4, pp. 178–180, Dec. 2002, https://doi.org/10.21091/mppa.2002.4028.

[4] L. De Smet, H. Ghyselen, and R. Lysens, “Incidence of overuse syndromes of the upper limb in young pianists and its correlation with hand size, hypermobility and playing habits,” *Ann. Chir. Main Memb. Super.*, vol. 17, no. 4, pp. 309–313, Jan. 1998, https://doi.org/10.1016/S0753-9053(98)80030-6.

[5] J. Farias *et al.*, “Anthropometrical analysis of the hand as a Repetitive Strain Injury (RSI) predictive method in pianists,” *Ital. J Anat. Embryol.*, vol. 107, no. 4, pp. 225–231, 2002.

[6] E. Booker and R. Boyle, “Piano keyboards–one size does not fit all! Pianistic health for the next generation.,” in *10th Austral. Piano Pedagogy Conf. Proc.*, Jul. 2011.

[7] CH. Wagner, “The pianist’s hand: anthropometry and biomechanics,” *Ergon.*, vol. 31, no. 1, pp. 97–131, Jan. 1988, https://doi.org/10.1080/00140138808966651.

[8] Y. Kaufman-Cohen, S. Portnoy, R. Sopher, L. Mashiach, L. Baruch-Halaf, and N. Z. Ratzon, “The correlation between upper extremity musculoskeletal symptoms and joint kinematics, playing habits and hand span during playing among piano students,” *PLOS ONE*, vol. 13, no. 12, p. e0208788, Dec. 2018, doi: 10.1371/journal.pone.0208788.

[9] S. Bruno, A. Lorusso, and N. L’Abbate, “Playing-related disabling musculoskeletal disorders in young and adult classical piano students,” *Int. Arch. Occup. Environ. Health.*, vol. 81, no. 7, pp. 855–860, Jul. 2008, https://doi.org/10.1007/s00420-007-0279-8.

[10] E. Yoshimura, P. M. Paul, C. Aerts, and K. Chesky, “Risk factors for piano-related pain among college students,” *Med. Probl. Perform. Art.*, vol. 21, no. 3, pp. 118–125, Sep. 2006, https://doi.org/10.21091/mppa.2006.3024.

[11] N. Sakai, M. C. Liu, F.-C. Su, A. T. Bishop, and K.-N. An, “Hand span and digital motion on the keyboard: Concerns of overuse syndrome in musicians,” *J. Hand Surg.*, vol. 31, no. 5, pp. 830–835, May 2006, https://doi.org/10.1016/j.jhsa.2006.02.009.

[12] K.-Y. Lai, S.-K. Wu, I.-M. Jou, H.-M. Hsu, M.-J. Chen Sea, and L.-C. Kuo, “Effects of hand span size and right-left hand side on the piano playing performances: Exploration of the potential risk factors with regard to piano-related musculoskeletal disorders,” *Int. J. Ind. Ergon.*, vol. 50, pp. 97–104, Nov. 2015, https://doi.org/10.1016/j.ergon.2015.09.011.

[13] R. Mailly, C. Turner, E. Goubault, F. Dal Maso, and F. Verdugo, “Impact of music expression-related parameters on pianists’ kinematics and muscle activity: A systematic review,” *Music Sci.,* vol. 8, 2025, https://doi.org/10.1177/20592043251317019.

[14] B. G. Wristen, M. C. Jung, A. K. G. Wismer, and M. S. Hallbeck, “Assessment of muscle activity and joint angles in small-handed pianists: A pilot study on the 7/8-sized keyboard versus the full-sized keyboard,” *Med. Probl. Perform. Art.*, vol. 21, no. 1, pp. 3–9, Mar. 2006, https://doi.org/10.21091/mppa.2006.1002.

[15] B. Wristen and M. S. Hallbeck, “The 7/8 piano keyboard: An attractive alternative for small-handed players,” *Update Appl. Res. Music Educ.*, vol. 28, no. 1, pp. 9–16, Nov. 2009, https://doi.org/10.1177/8755123309344162.

[16] J.-Y. Chi, M. Halaki, E. Booker, R. Boyle, and B. J. Ackermann, “Interaction between hand span and different sizes of keyboards on EMG activity in pianists: An observational study,” *Appl. Ergon.*, vol. 97, p. 103518, Nov. 2021, https://doi.org/10.1016/j.apergo.2021.103518.

[17] E. Yoshimura and K. Chesky, “The application of an ergonomically modified keyboard,” *MTNA e-journal*, pp. 2–13, 2009.

[18] C. Turner, E. Goubault, F. Dal Maso, M. Begon, and F. Verdugo, “The influence of proximal motor strategies on pianists’ upper-limb movement variability,” *Hum. Mov. Sci.*, vol. 90, p. 103110, Aug. 2023, https://doi.org/10.1016/j.humov.2023.103110.

Enhancement of IntuiSketch ITS with Teacher's Planned Sequence to Improve Learning by Drawing

Islam BARCHOUCH [1*], Swan SERRE [1*], Eric ANQUETIL [1*], Nathalie GIRARD [2*]
[1*] *Univ Rennes, INSA Rennes, CNRS, IRISA - UMR 6074, F-35000 Rennes, France.*
[2*] *Univ Rennes, CNRS, IRISA - UMR 6074, F-35000 Rennes, France.*
FirstName.LastName@irisa.fr

***Abstract*— IntuiSketch is a pen-based Intelligent Tutoring System (ITS) adapted to interpret semi-structured drawings. It supports learning by drawing and provides real-time feedback to students on pen-based tablets. This paper presents an enhancement of IntuiSketch, focusing on the integration of the teacher's pedagogical instructions into the Knowledge Graph (KG) used to evaluate the student's work. Previously, IntuiSketch evaluated student work based on domain knowledge only, independently of pedagogical intent. The aim of this paper is to introduce a two-level feedback mechanism that incorporates domain knowledge and the pedagogical steps defined by the teacher, enabling the system to evaluate both the accuracy and the pedagogical correctness of the student's sketches. When students draw a specific structure in the wrong step, the system generates deferred feedback to guide them to follow the planned sequence of the exercise. By integrating the pedagogical constraints assigned by the teacher, IntuiSketch incorporates both domain requirements and pedagogical objectives, promoting a complete learning experience. The system was tested on a spine exercise involving 13 student realizations (595 analyzed strokes), showing high recognition performance and relevant feedback generation.**

Keywords*— *Adaptive Learning, Online Sketch Recognition, Constraint-based ITS, Knowledge Graph

I. Introduction

Learning by drawing has been proven to be an effective pedagogical approach, particularly in complex fields such as anatomy [1] [2]. IntuiSketch is conceived to help students learn anatomy by providing real-time feedback on sketches drawn using pen-based tablets (Fig. 1(b)). Inspired by IntuiGeo architecture [3], an ITS for geometry learning, IntuiSketch's architecture is based on two main engines: the recognition engine and the supervision engine (see grey blocks in Fig. 2 for reference) [4]. The recognition engine (see sec. II. A.) is based on the Context-Driven Constraint Multiset Grammar (CD-CMG) formalism [5] combined with Fuzzy Logic [6] and an Incremental Classifier that learns from a few examples [7]. It interprets semi-structured sketches by applying shape, spatial and geometric constraints to evaluate their accuracy. The supervision engine (see sec. II. B.) then evaluates the recognized sketches based on a Knowledge Graph (KG), which incorporates domain-specific constraints and generates corrective feedback if constraints are not respected. IntuiSketch is able to guide students through an exercise organized by the teacher into different steps, each providing instructions and a reference drawing to help them accurately reproduce the required elements (Fig. 1). It uses two types of feedback: immediate feedback, which relies on primary constraints essential for eager interpretation, and deferred feedback, which uses secondary constraints to more finely evaluate students' sketches but are not involved in the recognition process [8][9].

The KG, generated dynamically based on the teacher's exercise, represents the core of IntuiSketch. It enables the system to evaluate what the student draws on the fly compared to the structure and constraints defined in the KG, providing real-time feedback to guide the student. However, the existing approach focuses on evaluating the accuracy of student work in relation to the domain knowledge represented by the KG, which is based entirely on the CD-CMG formalism defined manually by the developer, without taking into account adherence to the sequence or pedagogical intent defined by the teacher. This limitation allows students to complete the exercise in any sequence without respecting the steps defined by the teacher in the instructions, which may be in conflict with the pedagogical intention.

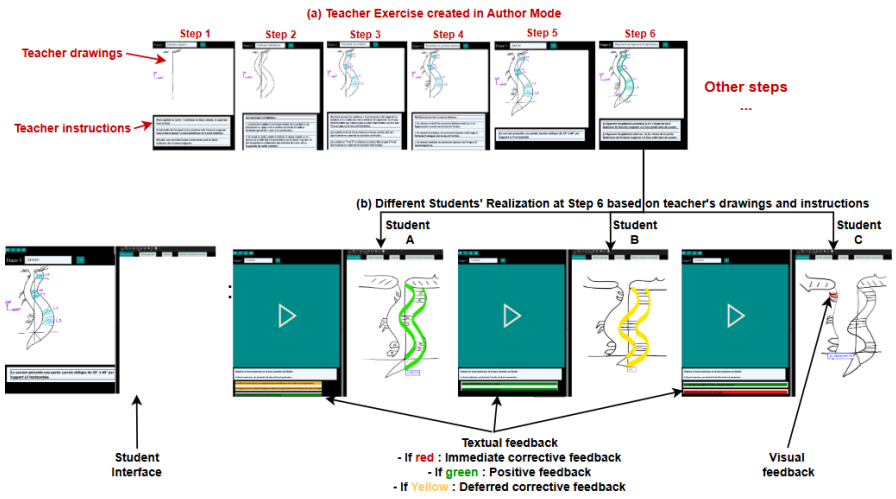


Fig. 1. Anatomy exercise organized into steps by the teacher and examples of different students' realizations at step 6 receiving immediate or deferred feedback

IGS 2025 ©2024 IEEE

To address this limitation, the proposed enhancement integrates the pedagogical instructions into the KG [10]. The system's new version includes the teacher's planned step sequence, which is now represented in the KG as pedagogical constraints. Thus, each node of the KG contains specific attributes related to the step of the exercise in which the element was drawn by the teacher. During the evaluation process, the system evaluates not only the structural accuracy of the element, but also its conformity with the pedagogical order. If a student draws an element in the wrong step, the system generates deferred feedback to guide the student in following the intended learning progression. This approach ensures that the system takes into account both domain-specific precision and pedagogical alignment, providing extensive feedback that improves the learning experience.

II. Existing IntuiSketch Architecture

A. Recognition Engine: Online Interpretation of Semi-Structured Sketches

IntuiSketch's recognition engine is based on the CD-CMG formalism [5] combined with Fuzzy Logic [6] and an Incremental Classifier [7]. This hybrid approach combines structural and statistical approaches. Using the shape, geometric and spatial constraints defined by CD-CMG rules, the formalism enables the system to interpret semi-structured sketches, which do not always have predefined geometrical structures, with a high degree of flexibility . Fig. 2 presents the interpretation workflow.

a) CD-CMG Formalism and Evolve Classifier: The CD-CMG rules have three stages: preconditions, constraints and postconditions. These stages guide the recognition process by defining the spatial, structural and contextual relationships between elements [5]. First, preconditions ensure that the system is able to manage spatial relationships between elements allowing it to evaluate whether an element is correctly placed in a defined zone, or whether it meets specific relationship requirements. Second, constraints ensure that geometric properties and shapes of elements are respected, making it possible to recognize elements that are structurally valid. Finally, postconditions dynamically update the drawing context, allowing the recognition engine to adapt to the evolution of the document and to recognize the following elements. To recognize semi-structured shapes that are complex and difficult to model using mathematical constraints, the Incremental Classifier Evolve extends CD-CMG capabilities with the ability to learn from a limited number of examples, enabling the system to recognize irregular shapes and adapt to different drawing styles.

b) Constraints Categorization (Primary and Secondary Constraints): To prevent repetitive interruptions of the student's workflow due to the generation of immediate feedback for every error, independently of its severity, and to introduce more flexibility into the recognition process, the constraints were categorized by extending the CD-CMG formalism as primary and secondary. Primary constraints represent essential requirements that must be satisfied for immediate recognition. These constraints focus on critical aspects of the sketch and ensure their accuracy. Secondary constraints, on the other hand, allow minor errors to be made, enabling the recognition process to continue without interruption. These constraints do not interfere in the recognition process, but help to guide students to improve their drawings later, without disrupting their creative workflow.

B. Supervision Engine: Knowledge Graph and Feedback Generation

IntuiSketch's supervision engine is in charge of evaluating student sketches compared to a structured representation of the exercise created by the teacher. This evaluation is based on a dynamic KG that helps to provide personalized feedback and guide students through the learning process. Fig. 2 presents the evaluation workflow.

a) Knowledge Graph Creation: The KG is generated dynamically when the student loads an exercise in student mode, based on what the teacher has drawn in author mode, which is the reference model. It serves as a structured representation of the exercise, containing the elements to be drawn, their properties and the relationships between them. In author mode, the teacher draws the elements step by step. Each stroke drawn by the teacher is analyzed and recognized using the recognition engine and the recognized elements therefore represent the exercise model. Once the student loads the exercise, the KG is initialized by associating each element defined by the teacher to a node in the graph. Its edges represent the constraints related to each element (node). Some constraints ensure the recognition of the element itself by imposing shape and geometric properties, while others define the spatial relationships between elements, capturing the dependencies that structure the overall exercise. As the student progresses through the exercise, each new stroke is analyzed and associated with a corresponding node in the KG. The constraints presented by the edges provide the basis to evaluate the accuracy of the student's sketches. If a match is found, the associated constraints are evaluated to generate corrective feedback if not respected, and the KG is updated to identify the student's progress and what remains to be evaluated in the current state of the exercise.

b) Feedback Categorization (Immediate and Deferred Feedback): The main objective of the supervision engine is to provide adaptive feedback that guides students through the exercise. Each feedback is categorized as immediate or deferred [8][9]. Immediate feedback is generated when primary constraints are not respected. This type of feedback ensures that critical errors, which could prevent progress, are addressed on the fly, helping to reinforce understanding of fundamental concepts. Deferred feedback, on the other hand, is applied when secondary constraints are not satisfied. It is generally provided at the end of a step or upon the student's request. Once important steps of the exercise are complete, the system provides qualitative deferred feedback for further refinement.

III. Contributions

Currently, the system focuses on analyzing students' strokes, but does not take pedagogical progress into account. To address the limitations related to the insufficient integration of pedagogical intent in the evaluation process, we propose to integrate pedagogical constraints directly into the knowledge graph (KG). These constraints are based on the teacher's exercise model and enable the system to assess not only the shape, geometric and spatial accuracy of a student's work, but also its conformity with the pedagogical sequence defined by the teacher. Contributions are represented in red in Fig. 2.

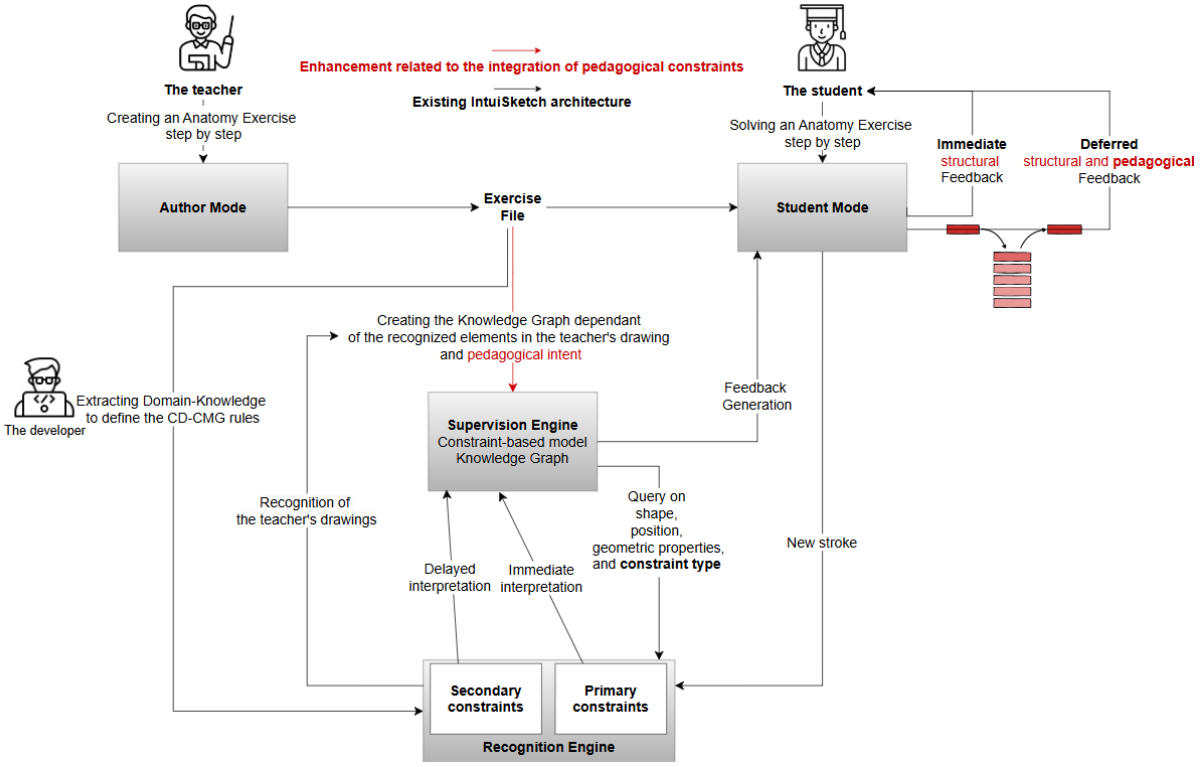


Fig. 2. IntuiSketch System Workflow: Implementation of Pedagogical Constraints

The exercise model created by the teacher is structured as a sequence of steps (Fig. 1(a)). Although each step contains a set of elements to be drawn following a sequence of instructions, IntuiSketch can still recognize and evaluate drawings on the basis of their structural and semantic accuracy. Within a step, elements do not have fixed sequence, so students can draw them in any order without affecting the overall structure. However, between steps, it is often necessary to follow the sequence required, as elements in a step are often identified based on those drawn in previous steps. When a student loads the exercise, the system processes this teacher model to dynamically generate the KG. The nodes corresponding to the recognized elements in the teacher's exercise are completed with attributes that indicate the specific step at which the element node was drawn. This ensures that the pedagogical structure of the exercise is preserved, and that each element represented in the KG reflects the sequence followed by the teacher when creating the exercise. The edges of the graph still represent constraints that define the shape, geometric, and spatial relationships of elements. However, the main contribution is the introduction of pedagogical constraints in the KG. These constraints specify the expected sequence of the exercise, enabling the system to evaluate the step in which an element has been drawn by a student against the step in which it is supposed to be drawn.

When a student starts an exercise, the teacher step's drawing model is first displayed, then hidden when the student begins drawing, to encourage memorization. The system then attempts to recognize the new stroke by verifying shape, geometric, and spatial constraints. If the element is not correctly drawn and cannot be recognized, immediate and/or deferred feedback is provided (Fig. 1). Deferred feedback now includes structural and pedagogical feedback. The system can evaluate whether an element has been drawn in the correct step defined by pedagogical constraints. This approach first validates the structural aspects of the drawing, allowing it to be recognized and enabling the student to correct errors related to the domain, and then checks whether students have respected the pedagogical sequence defined by the teacher, in order to improve their learning experience without interruption. To further support this approach, the system prevents students from accessing the next step until all required elements have been drawn. When an attempt is made to skip mandatory content, a warning is displayed. Additionally, if the student has not opened the tutorial, a reminder is shown at the end of the step to encourage adherence to the teacher's instructions.

IV. Results and Discussion

To validate the system's ability to recognize anatomical sketches and generate correct feedback in line with pedagogical constraints, a Spine exercise created by a teacher and structured in 9 steps (Fig. 1) was completed by students from the IRISA laboratory in Rennes. The main objective was to assess whether, for 13 student realizations, the system could accurately recognize the anatomical sketches at the correct step and generate feedback that guides the student both pedagogically and structurally. The test protocol includes student drawings saved after completion, which are then compared with the ground truth provided by the teacher. Table 1 presents the feedback generated by the system on **13 student realizations**, enabling a total of **595 strokes** to be obtained, while Table 2 provides a detailed analysis of the system's performance.

TABLE I. Summary of Test Database for System Evaluation

Data	***Feedback***	***Positive feedback***	***Immediate feedback***	***Deferred feedback***	***Pedagogical feedback***
Total Number	596	454	77	65	24

TABLE II. Performance Metrics for System Evaluation

Metrics	***Precision***	***Recall***	***Error rate***
Recognition	0.994	0.926	0.08
Feedback	0.879	0.937	0.167

Results show that the system provides a significant number of feedback to guide students, with a total of 596 feedback, including 454 positive feedback (i.e. well recognized elements), 77 immediate and 65 deferred corrective feedback, of which 24 are pedagogical. Despite the difficulties, the system is robust enough to accurately recognize student sketches, with a recognition precision of 0.994 and a recall of 0.926. Feedback generation also performs extremely well, with a precision of 0.879 and a recall of 0.937. The integration of secondary feedback, including pedagogical advice, is essential as students often deviate from the intended pedagogical sequence. This integration enables the system not only to guide students effectively, but also to improve recognition accuracy. Without this pedagogical feedback, recognition errors could propagate, affecting both subsequent sketches and feedback generation. Additional evaluations, such as questionnaires on perceived feedback usefulness, are planned.

V. Conclusion and Perspectives

This paper presented a significant enhancement to IntuiSketch by integrating pedagogical instructions into the KG. These advances help to generate deferred feedback related to pedagogical constraints representing teacher instructions that are not respected by students and improve the alignment of domain-specific knowledge accuracy with pedagogical intent, thus creating a more effective and comprehensive learning experience. Future work aims to enable the teacher and student to interact with the system continually in order to improve its capabilities. By improving interaction between the system and its users, teachers can actively contribute to building domain knowledge by adding or refining constraints defined using the CD-CMG formalism, and students can also benefit from more accurate and contextualized guidance, especially in complex exercises where they sometimes need to ask for help and not only wait to receive it automatically.

VI. Acknowledgment

This work is funded by the ANR (Project number 21-CE38-0009). We sincerely thank our collaborators from the LP3C laboratory: L. Leconte, A. Laine, and E. Jamet, for their valuable contributions particularly for ergonomic improvements and experiments.

References

[1] Ainsworth, S., Prain, V., & Tytler, R. (2011). Drawing to learn in science. Science, 333(6046), 1096-1097.

[2] Joewono, M., Karmaya, I. N. M., Wirata, G., Widianti, I. G. A., & Wardana, I. N. G. (2018). Drawing method can improve musculoskeletal anatomy comprehension in medical faculty student. Anatomy & cell biology, 51(1), 14-18.

[3] Krichen, O., Anquetil, E., & Girard, N. (2020). IntuiGeo: Interactive tutor for online geometry problems resolution on pen-based tablets. In ECAI 2020 (pp. 1842-1849). IOS Press.

[4] Barchouch, I., Krichen, O., Anquetil, E., & Girard, N. (2023, October). IntuiSketch, a pen-based tutoring system for anatomy sketch learning. In International Graphonomics Conference (pp. 61-74). Cham: Springer Nature Switzerland.

[5] Macé, S., & Anquetil, E. (2009). Eager interpretation of on-line hand-drawn structured documents: The dali methodology. Pattern Recognition, 42(12), 3202-3214.

[6] Zadeh, L. A. (1965). Fuzzy sets. Information and Control, 8(3), 338–353.

[7] Almaksour, A., & Anquetil, E. (2011). Improving premise structure in evolving takagi–sugeno neuro-fuzzy classifiers. Evolving Systems, 2, 25-33.

[8] Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. Frontiers in Psychology, 10, 3087.

[9] Shute, V. J. (2008). Focus on formative feedback. Review of Educational Research, 78(1), 153–189.

[10] Z. Liu, S. X. Yin, C. Lee and N. F. Chen, "Scaffolding Language Learning via Multi-modal Tutoring Systems with Pedagogical Instructions," 2024 IEEE Conference on Artificial Intelligence (CAI), Singapore, Singapore, 2024, pp. 1258-1265.

Hypothesis-Aware Ductus Generation for Greek Papyri: A Data-Driven Approach

Dinh Dung Van
University of Basel
Switzerland
dinhdung.van@unibas.ch

Giuseppe De Gregorio
University of Basel
Switzerland
giuseppe.degregorio@unibas.ch

Rodrigo C. G. Pena
Center for Data Analytics (CeDA)
University of Basel, Switzerland
rodrigo.cerqueiragonzalezpena@unibas.ch

Andreas Fischer
University of Fribourg and HES-SO
Switzerland
andreas.fischer@unifr.ch

Isabelle Marthot-Santaniello
University of Basel
Switzerland
i.marthot-santaniello@unibas.ch

***Abstract*— Greek papyri are essential for understanding Greco-Roman Egypt, offering insights into various aspects of life. Their handwritten scripts also provide a unique opportunity to explore the evolution of letterforms and the writing process. However, the degradation and fragmentation of these texts lead to multiple plausible writing trajectories, making traditional trajectory recovery methods insufficient. This study enhances conventional trajectory recovery by integrating expert knowledge. We first present the Alpha Dataset which includes recorded motions for Greek-Alpha ductus hypotheses, to examine letterform and writing hypothesis relationships. Second, we propose a hypothesis-aware ductus generation module to produce trajectories that visually align with papyrus characters while preserving ductus hypotheses. Our method serves as both an automated annotation tool for experts and a retrieval system for finding characters with similar stroke patterns across manuscripts, contributing to a deeper understanding of Greek papyrus writing dynamics.**

I. Introduction

Greek papyri represent an invaluable resource for understanding classical Antiquity. Due to the combination of papyrus as a writing material and the favorable climatic conditions of Egypt, a substantial number of these documents have been preserved across centuries to the present day [1]. These texts are of critical importance to papyrologists, historians, and philologists, as they provide direct insights into the administrative, economic, social, and cultural aspects of Greco-Roman Egypt. Beyond their content, papyri also hold considerable importance from a paleographic perspective [2]. The handwritten scripts preserved in these documents offer a unique opportunity to examine the graphic evolution of letters, as well as the movements and forms inherent in the writing process.

A fundamental challenge in the paleographic analysis of papyri is the absence of a well-defined ground truth [3]. Specifically, there is no direct access to the precise manner in which a letter or word was originally written. This uncertainty complicates the determination of the ductus, that is, the sequence, direction, and speed with which a script was executed. The ductus is a crucial element in paleographic studies, as it provides insights into the dynamics of writing, enabling scholars to formulate hypotheses regarding letter formation, stroke sequencing, and the execution speed of the scribe. Understanding the ductus is essential for the classification and comparative analysis of different manuscript traditions. However, in the absence of direct records of scribal movement, scholars must rely on inferential methods derived from the visual characteristics of the written traces. To address these challenges, computerized paleography tools have been developed, including handwriting analyzers and paleographic classification systems [4], [5]. However, the sparsity of representation in document analysis presents ongoing challenges for existing solutions.

This study seeks to address this issue by introducing an automated methodology for generating ductus hypotheses specifically for the Greek letter alpha, as its wide spectrum of calligraphic variation supports research into the evolution of handwriting and culture. To this end, a dedicated dataset, the Alpha-Dataset, has been constructed comprising motion recordings associated with different variants of alpha and the corresponding ductus hypotheses. Additionally, a hypothesis test dataset was created that consists of images of real alpha extracted from known papyri. For these characters, the ductus hypotheses were initially formulated through traditional paleographic analysis and subsequently compared with corresponding participant recorded writing movements. Finally, a model for ductus generation is proposed that aims to generate the trajectory of papyrus characters while preserving the ductus hypothesis from the Alpha-Dataset. This approach facilitates the generation of ductus trajectories by considering expert hypotheses, thereby contributing to a more nuanced understanding of writing dynamics in Greek papyri.

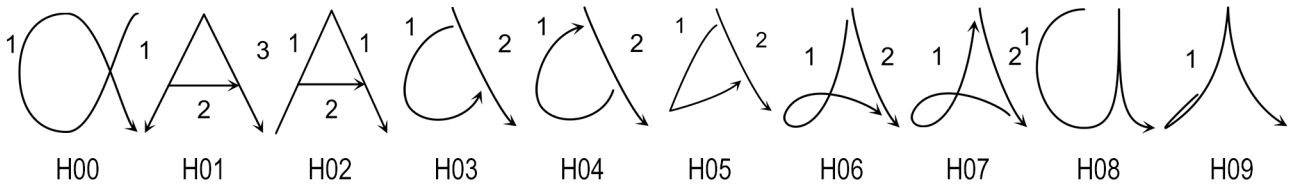


Fig. 1. The figure shows all ductus hypotheses recorded in the Alpha-Dataset. The number indicates the index of each stroke, which was defined by a pair of pen-up / pen-down movements.

II. The Datasets

A. The Alpha-Dataset

The Alpha-Dataset was constructed using the Wacom IoT tablet, a device that simulates handwriting on paper while allowing a guide template to be displayed on the screen. This setup enabled participants to follow an execution example and adhere to a predefined sequence of movements when tracing letters. A total of 10 ductus hypotheses were defined for different alpha variants, with each hypothesis recorded 10 times per user. The dataset comprises data from 57 participants, resulting in a total of 5,700 recorded writing instances of various alpha forms. In addition, the metadata consists of left/right-handed, the first script learned, and educational level, was collected for further analysis. Figure 1 illustrates all recorded hypotheses of the ductus as execution samples shown to participants. It should be noted that some hypotheses exhibit strong similarities, differing primarily in the sequence of strokes or the direction of execution (for example, the pairs H01-H02, H03-H04, and H06-H07).

To adapt users to tablet use and to ensure greater precision in data interpretation, four preliminary activities were recorded:

- Writing the sentence ”The quick brown fox jumps over the lazy dog” in the most natural way possible;
- Writing the letter α three consecutive times in the most spontaneous way for the subject;
- Writing the letter A three consecutive times in the most spontaneous way for the subject;
- Writing the letter *a* three consecutive times in the most spontaneous way for the subject.

After these adaptation activities, each subject completed ten repetitions for each of the ductus hypotheses, in random order to avoid learning effects. Each element of the data set consists of an .inkml file and a .csv file, containing detailed information about the recorded trace, including X and Y coordinates, pressure, the height of the pen tip from the paper, X and Y inclination of the pen, and time.

B. The Hypothesis Test Dataset

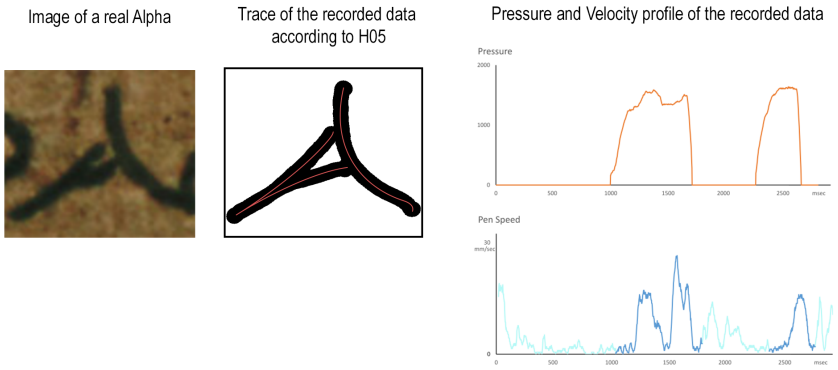


Fig. 2. Example of the test data. The figure shows an image of an alpha extracted from a papyrus, its trace recorded by an expert, and the related pressure and speed profiles.

The dataset includes images of the letter alpha, carefully selected by expert papyrologists and paleographers. These images come from real papyri, ensuring a high level of representativeness. For each alpha ductus hypothesis, ten representative images were selected for experts to record their retracing movements. An exception is the H00 hypothesis, for which only one image was identified, given its rarity in papyrus finds. The selected images were used as a reference to generate writing traces using the IoT tablet. The recording was performed according to the specific hypothesis of the ductus. During data acquisition, the

user was presented with the image of the character to be reproduced together with a black and white trace. Users were asked to retrace the letter, following the movement suggested by the indicated ductus hypothesis. Data were collected from five different users, each of whom performed one repetition for each image. In total, 455 writing traces were collected.

III. Hypothesis-Aware Ductus Generation Model

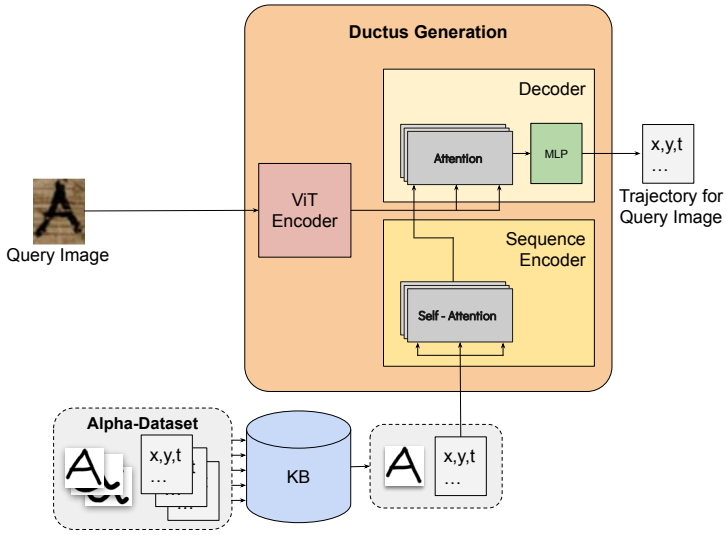


Fig. 3. The architecture of the ductus generation module. The module takes as input an image of the Greek letter alpha and a hypothesis recorded in the Alpha-Dataset stored in the Knowledge-base (KB). As output, it generates the trajectory on the query image.

Reconstruction of writing trajectories in ancient manuscripts presents significant challenges due to the inherent uncertainty of writing dynamics, the degradation of materials, and the fragmentation of characters. Ancient handwriting often exhibits multiple plausible trajectories, making conventional trajectory recovery methods inadequate. These variations cannot be reliably inferred from images alone and typically require expert domain knowledge to reconstruct the most likely writing path. To address this issue, we propose a hypothesis-aware ductus generation model, designed to integrate expert-driven hypotheses into trajectory generation, while preserving the structural integrity of possible ductus variations.

The core objective of our model is to align a ductus hypothesis trajectory with a query image of an alpha character written on papyrus. To achieve this, we employ an encoder-decoder architecture, as in Figure 3, consisting of three key components:

- **Vision Encoder**
  The vision encoder aims to extract spatial and structural features from the input papyrus image, to facilitate the generation of a hypothesis ductus given the target visual output. We utilize a Vision Transformer (ViT) [6] [7] to allow the model to localize key structural elements of the character, improving the alignment between the visual features and the generated trajectory.
- **Sequence Encoder**
  The sequence encoder aims to model the sequential dependencies of speed and position with respect to the intended input pattern. In order to facilitate the learning of these dependencies, we employ a self-supervised contrastive learning strategy as described by SIMCLR [8] on the self-attention mechanism [9]. For each sample from the Alpha dataset, we create two augmented samples that share a similar pattern but vary in execution speed at the stroke level, similar to the strategy described in [10]. By learning to extract similar features between these augmented samples, while differentiating from other input patterns, the sequence encoder can effectively model the intended input pattern. At a later stage, the sequence encoder can also be used as a discriminator to verify the pattern preservation during the generation process.
- **Decoder**
  The decoder aims to project the user input pattern onto a trajectory that visually aligns with the target image, by employing attention layers between visual and sequence features. Given a pair of augmented samples described in the section above, the decoder learns to generate one sample using pattern-related features extracted from the other sample in the same augmented pair. Consequently, the decoder is trained to generate a trajectory that is most similar to the intended input pattern, which focuses on handling the trade-off between visual alignment and input pattern preservation.

IV. Experimental Results

For each sample in the Alpha dataset, we resample the trace at a rate of 100 Hz and generate a pair of augmentations using a strategy similar to that described in [10]. The augmentation slightly varies the execution speed at the stroke level, introducing minor visual distortions while preserving the overall pattern. These patterns are captured by the Sequence Encoder, while the relationship between visual distortion and augmented execution speed is modeled by the Decoder, using vision features extracted by the Vision Encoder.

We train the model in a self-supervised manner, leveraging this approach to enhance robustness to noisy labels and improve generalizability to other character patterns. Representation learning is performed on the Sequence Encoder following the SIMCLR framework [8], encouraging the model to extract similar features for similar patterns and distinguish features across dissimilar ones. The Vision Encoder is jointly trained with the Decoder to learn the relations between visual distortion and augmented execution speed, conditioned by the pattern (or hypothesis) extracted by the Sequence Encoder. The Decoder outputs trajectories that optimize both pattern preservation and visual alignment constraints.

TABLE I. Alignment capability of using Alpha-Dataset hypothesis on Hypothesis Test Dataset images

Type	$LDTW_{prd}$	$AIoU_{prd}$	$LDTW_{org}$	$AIoU_{org}$
H00	5.41 ±0.89	0.33 ±0.027	38.56 ±4.69	0.08 ±0.031
H01	5.43 ±1.42	0.46 ±0.037	34.11 ±7.51	0.07 ±0.038
H02	5.06 ±1.09	0.45 ±0.033	44.20 ±6.00	0.04 ±0.024
H03	7.78 ±2.47	0.41 ±0.040	34.42 ±6.91	0.08 ±0.033
H04	6.99 ±2.41	0.41 ±0.039	38.19 ±6.43	0.06 ±0.032
H05	6.30 ±1.89	0.45 ±0.073	37.93 ±8.00	0.05 ±0.032
H06	5.91 ±1.33	0.45 ±0.060	43.19 ±8.22	0.04 ±0.026
H07	6.53 ±1.47	0.45 ±0.051	43.90 ±9.48	0.04 ±0.032
H08	7.41 ±2.39	0.37 ±0.041	33.86 ±4.74	0.07 ±0.026
H09	6.42 ±2.62	0.35 ±0.051	42.35 ±6.49	0.03 ±0.032
Validation Set	**6.33 ±2.11**	**0.42 ±0.059**	**39.27 ±8.26**	**0.06 ±0.035**

We denoted the evaluation result for the generated trajectories as ($LDTW_{prd}$, $AIoU_{prd}$) and for original hypotheses as ($LDTW_{org}$, $AIoU_{org}$). The prefix $_{org}$ indicates the evaluation using original input traces, while the prefix $_{prd}$ indicates the evaluation using generated traces.

To assess the model's ability to generate writing trajectories, we evaluate its performance using two complementary metrics. Length-independent Dynamic Time Warping (LDTW), which measures the similarity between the generated trajectory and the reference sequence, and Adaptive Intersection-over-Union (AIoU) [11], which evaluates visual alignment between the generated trajectory and the character image.

The model was trained on the Alpha-Dataset, using its recorded trajectories as input hypotheses. To evaluate alignment capability, we tested the model on the Hypothesis Test Dataset, comparing the generated trajectories with expert-annotated patterns from real papyrus images.

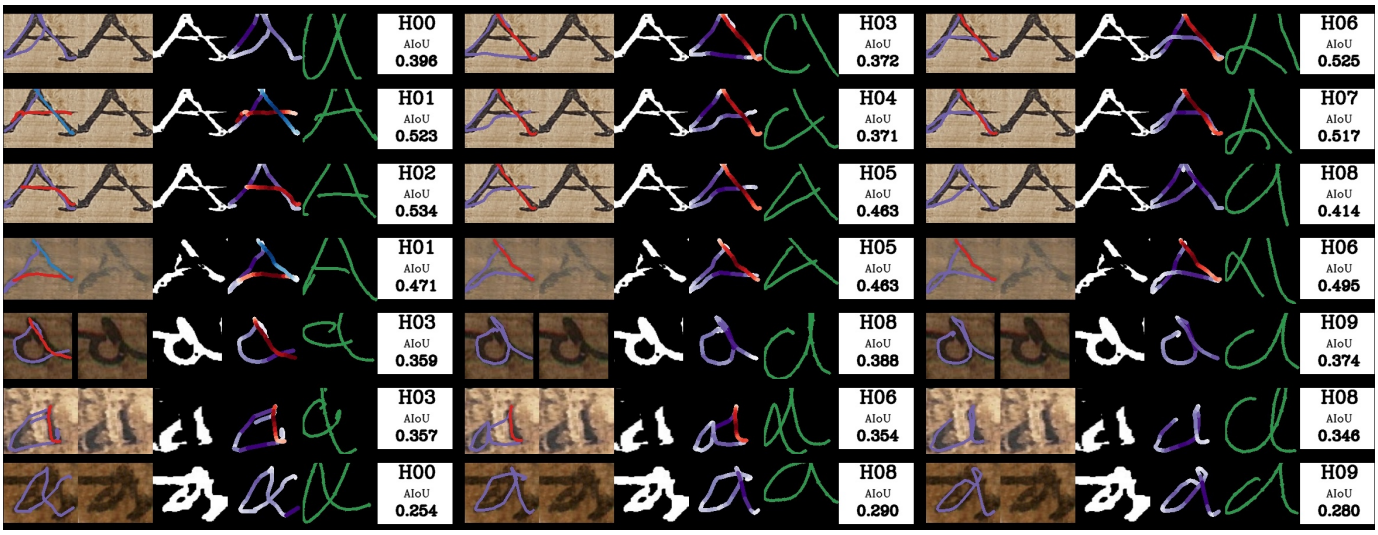


Fig. 4. Qualitative results showing the alignment of trajectories from the Alpha-Dataset with alpha images extracted from real papyri. From left to right: comparison between reconstructed trajectory and input image, original papyrus image, binarized papyrus, output trajectories with stroke-wise color coding, and input hypothesis in green strokes. In the fourth columns, each stroke is represented by a color, where the color density represent strength of relative speed.

Figure 4 presents qualitative alignment results in some unlabeled papyrus images. The first three rows present multiple attempts to align different hypotheses to the same alpha image. By ranking the AIoU metric, we observe that some hypotheses exhibit a higher probability of representing the true trajectory than others. In the fourth row, we can notice that the model successfully aligns multiple hypotheses (H01: 0.471, H05: 0.463, H06: 0.495), reinforcing the idea that multiple valid ductus reconstructions can exist for a single letterform. The last three rows show the model's ability to handle complex or degraded papyrus patterns. Although the results may exhibit local misalignments due to the retention of unwanted characteristics from the input hypotheses (i.e. small cursive or undesirable relative position constraints), the model effectively produces trajectories that fit the global context preserving the characteristics of the input ductus hypothesis.

V. Conclusions

In this study, we introduced a model for reconstructing the trajectory of alpha characters by aligning the recorded ductus hypotheses with the observed letterforms. Through the construction of the Alpha-Dataset, which captures motion recordings of alpha variants, and a validation dataset based on real papyrus samples, we established a computational framework that integrates traditional paleographic expertise with machine learning. The results demonstrate the capability of the model to reconstruct character trajectories while preserving the structural integrity of ductus hypotheses. The ability to align multiple plausible trajectories to a single letterform underscores the importance of hypothesis-aware reconstruction, particularly when dealing with uncertain or degraded ancient scripts. Some hypotheses, such as H01–H02, H03–H04, and H06–H07, exhibit indistinguishable offline shapes, suggesting that rather than determining a single definitive trajectory, the system could propose multiple plausible stroke sequences. This capability offers scholars a broader perspective on scribal variations and handwriting reconstruction. Although the model exhibits an adaptability to noise, its sensitivity to stroke positioning suggests areas for future improvement. Including additional contextual information or refining the alignment process could enhance trajectory consistency across different query images. Despite these challenges, our approach represents a step forward in computational paleography, providing a structured method for ductus inference. Future work will extend this methodology to additional Greek characters and enhance generalization to real-world papyrus images. By combining historical expertise with modern computational techniques, this research contributes to a deeper understanding of ancient writing practices and their evolution.

VI. Acknowledgments

This work has been funded by the Swiss National Science Foundation, SNSF Starting Grant TMSGI1-211682 "EGRAPSA : Retracing the evolutions of handwritings in Greco-Roman Egypt thanks to digital paleography" (Basel, 2023-2028)

References

[1] I. S. Moyer, "Greek and latin sources," in *The Oxford Handbook of Egyptology*. Oxford University Press, 11 2020. [Online]. Available: https://doi.org/10.1093/oxfordhb/9780199271870.013.67

[2] G. Cavallo, "Greek and latin writing in the papyri," in *The Oxford Handbook of Papyrology*. Oxford University Press, 12 2011. [Online]. Available: https://doi.org/10.1093/oxfordhb/9780199843695.013.0005

[3] D. N. Bonilla, "Training eyes and training hands in the digital research with manuscripts," *Open Information Science*, vol. 5, pp. 1–10, 06 2021.

[4] H. Mohammed, "Handwriting analysis tool v3.5 (hat3.5)," May 2021. [Online]. Available: https://doi.org/10.25592/uhhfdm.9179

[5] E. Werner and H. Mohammed, "Visual similarity annotation tool (v-sat)," Sep. 2024. [Online]. Available: https://doi.org/10.25592/uhhfdm.14851

[6] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2021. [Online]. Available: https://arxiv.org/abs/2010.11929

[7] L. Beyer, X. Zhai, and A. Kolesnikov, "Better plain vit baselines for imagenet-1k," 2022. [Online]. Available: https://arxiv.org/abs/2205.01580

[8] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," 2020. [Online]. Available: https://arxiv.org/abs/2002.05709

[9] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092523122100477X

[10] A. Fischer, R. Plamondon, C. O'Reilly, and Y. Savaria, "Neuromuscular representation and synthetic generation of handwritten whiteboard notes," in *2014 14th International Conference on Frontiers in Handwriting Recognition*, 2014, pp. 222–227.

[11] Z. Chen, D. Yang, J. Liang, X. Liu, Y. Wang, Z. Peng, and S. Huang, "Complex handwriting trajectory recovery: Evaluation metrics and algorithm," in *Computer Vision – ACCV 2022*, L. Wang, J. Gall, T.-J. Chin, I. Sato, and R. Chellappa, Eds. Cham: Springer Nature Switzerland, 2023, pp. 58–74.

A Dual-Backbone Framework for Alzheimer's Detection through Handwriting Multiple Sources

Cesare Davide Pace, Emanuele Nardone, Tiziana D'Alessandro
DIEI, University of Cassino and Southern Lazio, Via G. Di Biasio 43, 03043 Cassino (FR), Italy

***Abstract*— In the field of neurodiagnostics, the early detection of Alzheimer's disease remains a critical challenge, requiring non-invasive and scalable approaches. Handwriting analysis has demonstrated potential in capturing neuromotor impairments associated with Alzheimer's disease, as handwriting deterioration provides an early indicator of cognitive decline. This study proposes a novel deep-learning framework that integrates offline and RGB handwriting images to assist Alzheimer's disease detection. The offline images capture the static handwritten trace, while the RGB modality includes in-air traits encoding handwriting dynamics derived from raw data, thereby enriching feature representation. A dual-backbone Convolutional Neural Network backbone processes these modalities separately, followed by adaptive pooling, layer normalisation, and feature fusion. A Convolutional Block Attention Module also refines feature representations before final classification via a Multi-Layer Perceptron. While including synthetic RGB data provides a minor performance improvement, our findings indicate that the real offline images of the handwriting tasks encapsulate the most relevant information, preserving the morphological characteristics and distinctive neuromotor traits essential for Alzheimer's disease diagnosis. Experimental results confirm that our framework outperforms task-specific approaches, reinforcing the effectiveness of integrating multiple handwriting tasks into a unified model. Our previous findings demonstrated that consolidating all tasks yields superior diagnostic performance, and the proposed system further refines this strategy. This approach remains a robust and cost-effective solution for early Alzheimer's disease screening, with substantial implications for clinical applications and early diagnostics.

Keywords*— *Medical Expert Systems, Cognitive Impairments, Handwriting Analysis, Convolutional Neural Network.

I. Introduction

The incidence of neurodegenerative diseases (NDs) has markedly increased, primarily driven by global demographic shifts toward ageing populations [1]. Among these disorders, Alzheimer's disease (AD) is particularly concerning, as it leads to severe cognitive impairments—including memory loss, language deficits, and visuospatial dysfunction accompanied by a gradual decline in fine motor skills [2], [3]. Early diagnosis is crucial, given that current treatments primarily alleviate symptoms rather than halt disease progression, while emerging disease-modifying therapies are most effective in the earliest stages [4]. As a result, the scientific community is increasingly focused on identifying non-invasive biomarkers capable of detecting AD in its prodromal phase. One promising approach is handwriting analysis, which has demonstrated sensitivity to both neuromotor and cognitive decline [5], [6]. Handwriting, a complex sensorimotor activity, demands precise coordination between cognitive planning and motor execution. AD disrupts this interplay, often resulting in micrographia, irregular stroke kinematics, and spatial disorganisation [1], [7]. While tablet-based dynamic feature analysis has been prominent in computational studies, particularly for Parkinson's disease (PD) [7], AD research has faced limitations due to small, imbalanced datasets and task-specific modelling approaches [8], [9]. Recent advancements have shown that offline handwriting images, captured through standardised paper-and-pen tasks and digitised via high-resolution scanning, preserve diagnostically relevant spatial, morphological and pressure patterns while simplifying data acquisition [10], [11]. This approach aligns with clinical workflows, facilitating retrospective analysis of historical documents and enabling broader population screening [12].

Early studies on AD detection through handwriting analysis utilised machine learning on handcrafted features from individual tasks, yielding moderate accuracy but limited generalizability [13], [14]. Subsequent work explored deep learning with synthetic images derived from online handwriting dynamics, highlighting the potential of convolutional neural networks (CNNs) in capturing AD-specific artefacts [11], [12]. A pivotal breakthrough was achieved with offline image analysis, where pre-trained CNNs attained notable accuracy by learning latent representations directly from scanned handwriting samples [10].

Building on these findings, our recent study [15] introduced a unified model for all offline handwriting image tasks. In the current work, we extend this approach by processing offline handwriting images, which capture the static handwritten trace, and RGB representations encode dynamic handwriting information. Our methodology utilises two CNN backbones, which extract features independently, followed by feature fusion and classification. This innovative methodology achieves results by focusing on the real morphological handwriting shape from offline images and capturing its dynamic features, including in-air traits, from RGB images. This combined approach improves diagnostic accuracy and provides a robust framework for early AD detection.

II. Materials

The proposed study considers two types of handwriting images: offline and synthetic in-air-on-paper RGB images derived from the same samples collected using a WACOM Bamboo Folio tablet and a sensorized pen. Such a tool captured spatial coordinates (x, y), pressure (z), and timestamps at 200Hz for movements within 3 cm from the tablet surface. Moreover, the graphic tablet had paper sheets fastened on its surface, guaranteeing that the real handwriting traces were also stored. Each

TABLE I. AVERAGE (STD) DEMOGRAPHIC DATA OF PARTICIPANTS.

	Age	Education	#Women	#Men
AD patients	71.5 (9.5)	10.8 (5.1)	46	44
HC	68.9 (12)	12.9 (4.4)	51	39

dataset (offline and RGB) consists of approximately 4350 images collected from 174 participants performing 25 handwriting tasks [16], designed with clinical experts to assess cognitive, memory and motor skills affected by AD. The tasks are divided into three categories: graphic, memory and dictation, and copy and reverse copy tasks, as detailed in Table II. Participants belonged to two classes, 85 healthy controls (HC) and 89 AD patients (AD), enrolled to match year, gender and education distribution, as highlighted in Table I. The acquisition followed strict recruiting criteria with the help of the Alzheimer's unit of the "Federico II" hospital in Naples. Every participant's cognitive status was assessed through standard tests such as the MMSE [17], FAB [18], and MoCA [19].

Offline images are direct digital representations of handwriting captured from traditional pen-and-paper tasks. These images are obtained by segmenting the scanned handwriting traces produced by participants on the paper sheets fixed on the tablet. This process guarantees real trace preservation. The captured handwriting reflects the actual strokes executed, preserving their natural curves and pressure variations. Additionally, these images are grayscale, meaning that they have only one colour channel, and the intensity of each pixel corresponds to the ink's natural shades, which depend on the pressure exerted by the writer and the dynamics of his movement. In-air-on-paper RGB images are synthetic handwriting reconstructions generated from online data. Unlike offline images, which preserve the natural ink trace on paper, these images reconstruct the handwriting trace by interpolating consecutive points. Each stroke encodes dynamic features in RGB channels: pressure (red), velocity (green), and jerk (blue), capturing both shape approximation and kinematic properties. Additionally, they include in-air movements made within 3 cm of the tablet surface, distinguishing them from offline images.

Further details about the materials considered for this study can be found in [10]–[12], [15].

III. METHODS

This study investigates the classification of handwriting data for AD diagnosis using CNN-based approaches. Specifically, we process two different sources of handwriting data, described in Section II: offline handwriting and in-air-on-paper RGB images. We evaluate multiple model configurations based on different feature fusion strategies, attention mechanisms, and backbone architectures to determine the optimal approach for distinguishing between HC and AD.

A. Dual-Backbone architecture

The proposed dual-backbone architecture comprises two independent backbones. Thus, one backbone is responsible for extracting features from offline handwriting images, which capture static representations of handwritten images. The second

TABLE II. LIST OF THE TASKS: MEMORY AND DICTATION (M), GRAPHIC (G), AND COPY (C).

Task	Description	Category
1	Signature	M
2	Join two points with a horizontal line, continuously four times	G
3	Join two points with a vertical line, continuously four times	G
4	Retrace a circle (6 cm of diameter) continuously four times	G
5	Retrace a circle (3 cm of diameter) continuously four times	G
6	Copy the letters 'l', 'm', and 'p'	C
7	Copy the letters on the adjacent rows	C
8	Write a sequence of four lowercase letters 'l' cursive	C
9	Write a sequence of four lowercase cursive bigram 'le' cursively	C
10	Copy the word "foglio"	C
11	Copy the word "foglio" above a line	C
12	Copy the word "mamma"	C
13	Copy the word "mamma" above a line	C
14	Memorise the words "telefono", "cane", and "negozio" and rewrite them	M
15	Copy in reverse the word "bottiglia"	C
16	Copy in reverse the word "casa"	C
17	Copy six words (regular, non-regular, non-words) in the appropriate boxes	C
18	Write the name of the object shown in a picture (a chair)	M
19	Copy the fields of a postal order	C
20	Write a simple sentence under dictation	M
21	Retrace a complex form	G
22	Copy a telephone number	C
23	Write a telephone number under dictation	M
24	Draw a clock, with all hours and put hands at 11:05 (Clock Drawing Test)	G
25	Copy a paragraph	C

backbone processes in-air-on-paper RGB handwriting images that encode the dynamic characteristics of the writing process, both for the on-paper and in-air traits.

To effectively integrate information from these two sources, we explore different feature fusion strategies, which determine how the output feature vectors of the two backbones are combined before final classification. The key aspects of this fusion process are outlined below. The architecture is shown in Figure 1.

B. Feature Fusion Methods

To merge the extracted features from the two backbones, we investigate two primary strategies:

- **Concatenation**: The feature maps are concatenated along the channel dimension, preserving all extracted information while allowing the subsequent layers to learn joint representations of the two handwriting modalities.
- **Element-wise summation**: Feature maps from the two backbones are summed element-wise. This forces the model to integrate information at the feature level.

C. Feature Fusion Position

We also examine where in the network the feature fusion occurs using two possible configurations:

- **Early fusion**: The feature maps extracted by both backbones are merged immediately after the backbone output before passing through any attention modules. Early fusion enables early-stage interaction between offline and in-air-on-paper handwriting features, allowing the network to refine a unified representation.
- **Late fusion**: In this setup, each backbone independently processes the handwriting images through its respective attention module before the features are merged later. This method ensures that each modality undergoes its refinement before being integrated.

D. Attention mechanism

Before the classification part, we incorporate attention mechanisms to improve the model's ability to focus on the most relevant features. We explore two attention modules:

- **Convolutional Block Attention Module (CBAM) [20]**: This module applies both spatial attention and channel-wise attention, ensuring that the network focuses on key information in both handwriting image modalities.
- **Self-Attention [21]**: Self-attention enables the model to capture long-range dependencies by computing relationships between all positions in the feature maps. This allows the network to dynamically emphasise relevant handwriting features regardless of their spatial location, making it particularly effective for capturing global structural patterns in handwriting.

E. 4-channel Input Architecture

As an alternative to using separate backbones, we explore a 4-channel input model by modifying the first convolutional layer of the backbone. Since offline handwriting images are grayscale (1 channel) and in-air-on-paper handwriting images are RGB (3 channels), we concatenate them into a single 4-channel input. This approach allows a single backbone to process both handwriting images directly. We investigate different attention mechanisms and fusion methods within this architecture, as with the Dual-Backbone architecture.

F. Single-Backbone Baseline

We also compare single-backbone models trained on only one handwriting image group (offline or in-air-on-paper). These serve as baselines to measure the benefit of multimodal integration. All experiments employ the ConvNeXtSmall [22] architecture as the feature extraction backbone. This decision is based on prior research in handwriting analysis [15], where ConvNeXtSmall demonstrated superior performance over other Convolutional and Transformer backbones for handwriting recognition.

TABLE III. PERFORMANCE COMPARISON OF DIFFERENT MODEL CONFIGURATIONS.

Method	Attention	Fus. Method	Fus. Position	Accuracy	F1 Score	Sensitivity	Specificity	Precision	MCC
Dual	CBAM	Concat	Early	79.9 (4.2)	80.6 (4.9)	80.8 (9.7)	79.1 (3.9)	81.0 (4.1)	60.1 (8.2)
Single (offline)	x	x	x	78.9 (4.3)	80.3 (4.3)	82.6 (8.7)	75.4 (2.8)	78.7 (4.0)	58.1 (9.0)
4-channel	CBAM	x	x	72.4 (7.6)	71.8 (8.2)	67.4 (8.7)	77.8 (7.0)	77.0 (8.1)	45.2 (15.2)
Single (RGB)	x	x	x	66.8 (5.8)	66.8 (7.7)	65.3 (14.6)	68.7 (12.2)	70.3 (7.3)	34.7 (10.7)

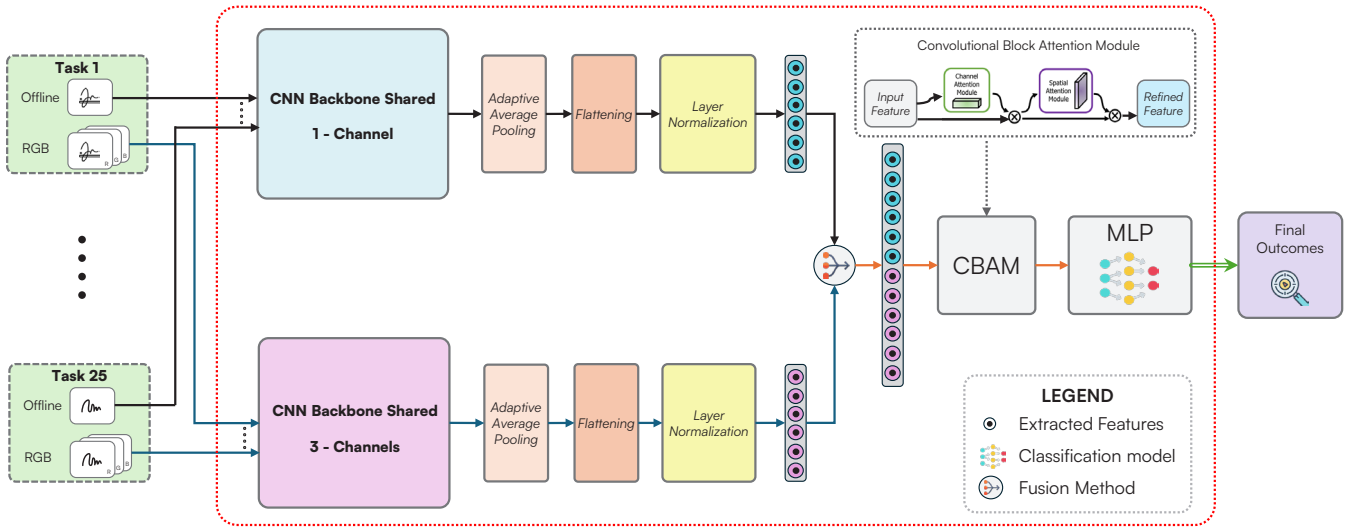


Fig. 1. Scheme of the proposed approach.

IV. Results

Table III presents the performance comparison, based on key classification metrics, of the experimental approaches detailed in Section III. Results are reported through key classification metrics, including accuracy, F1-score, sensitivity, specificity, precision, and Matthews correlation coefficient (MCC), each computed as the mean and standard deviation over multiple folds, as each experiment was executed by applying a 5-fold cross-validation strategy.

The results demonstrate that the dual-backbone CNN with CBAM attention and early fusion is the most effective approach for handwriting-based AD detection, achieving the highest overall accuracy (79.9%) and a strong balance between sensitivity (80.8%) and specificity (79.1%). This suggests that integrating information from both in-air-on-paper RGB and offline handwriting representations in a structured manner improves diagnostic performance. The single-backbone model fed with offline images only performs comparably (78.9%) but prioritises sensitivity (82.6%) over specificity (75.4%), suggesting it is better suited for cases where early detection is critical. In contrast, the 4-channel input model underperforms (72.4%) due to its inability to effectively integrate offline and in-air-on-paper RGB information within a single backbone. Moreover, this approach shows struggles with sensitivity (67.4%), suggesting that a single backbone may not effectively extract complementary information from both image groups. Finally, the single-backbone model fed through in-air-on-paper RGB images is the least effective (66.8%), confirming that offline images contain the most relevant diagnostic features.

V. Conclusions

In this study, we explored the integration of deep learning techniques for handwriting-based AD detection, considering both offline grayscale images and synthetic in-air-on-paper RGB representations. Our findings highlight the critical role of structured feature fusion in maximising the diagnostic potential of multimodal handwriting data. These obtained results confirm that considering in-air-on-paper RGB and offline handwriting images can be beneficial, but only when the information from the two sources is fused in a structured and controlled manner. The dual-backbone approach with early fusion demonstrates this advantage, achieving the highest accuracy by effectively integrating the complementary characteristics of both image types. In contrast, the 4-channel single-backbone model performs worse, indicating that simply merging greyscale and RGB channels within a single network is insufficient. This suggests that unstructured integration may introduce redundant or conflicting information, reducing the model's ability to extract meaningful neuromotor traits.

Furthermore, the single-backbone offline model performs comparably to the best dual-backbone model, reinforcing the effectiveness of offline handwriting data. Despite RGB images incorporating both on-paper and in-air dynamic information and traits, they perform worse than offline, highlighting that the real morphological handwritten trace carries more significant neuromotor information than RGB images, where the trace is rebuilt and not the original. The superior performance of offline handwriting suggests that the static trace alone provides a more stable and diagnostically relevant representation of handwriting impairments in AD. In contrast, the additional dynamic cues in RGB may introduce noise or less reliable indicators. These findings suggest that while multimodal approaches offer promise, the offline handwriting trace remains the most stable and informative source for detecting handwriting abnormalities linked to AD. Future research should focus on refining fusion

techniques to maximise complementary information while mitigating redundancy. Additionally, integrating domain knowledge from neurodegenerative disease studies may further increase model interpretability and clinical applicability.

References

[1] G. Vessio, "Dynamic handwriting analysis for neurodegenerative disease assessment: A literary review." *Applied Sciences*, vol. 9(21):4666, 2019.

[2] J. Lambert, B. Giffard, F. Nore, V. de la Sayette, F. Pasquier, and F. Eustache, "Central and peripheral agraphia in alzheimer's disease: From the case of auguste d. to a cognitive neuropsychology approach," *Cortex*, vol. 43, no. 7, pp. 935–951, 2007.

[3] J. Neils-Strunjas, K. Groves-Wright, P. Mashima, and S. Harnish, "Dysgraphia in Alzheimer's disease: a review for clinical and research purposes," *J Speech Lang Hear Res*, vol. 49, no. 6, pp. 1313–30, 2006.

[4] J. A. R. Nicoll, G. R. Buckland, C. H. Harrison, A. Page, S. Harris, S. Love, J. W. Neal, C. Holmes, and D. Boche, "Persistent neuropathological effects 14 years following amyloid-β immunization in Alzheimer's disease," *Brain*, vol. 142, no. 7, pp. 2113–2126, 06 2019.

[5] C. De Stefano, F. Fontanella, D. Impedovo, G. Pirlo, and A. Scotto di Freca A, "Handwriting analysis to support neurodegenerative diseases diagnosis: a review," *Pattern Recognition Letters*, vol. 121, pp. 37–45, 2018.

[6] E. R. Kandel, J. H. Schwartz, and T. M. Jessell, *Principles of Neural Science*, 4th ed. McGraw-Hill Medical, Jul 2000.

[7] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Evaluation of handwriting kinematics and pressure for differential diagnosis of parkinson's disease," *Artificial Intelligence in Medicine*, vol. 67, pp. 39–46, 2016.

[8] P. Werner, S. Rosenblum, and A. Korczyn, "Handwriting process variables discriminating mild alzheimer's disease and mild cognitive impairment," *Journal of Gerontology: PSYCHOLOGICAL SCIENCES*, vol. 61, no. 4, pp. 228–36, 2006.

[9] P. Werner, S. Rosenblum, G. Bar-On, J. Heinik, and A. Korczyn, "Handwriting process variables discriminating mild alzheimer's disease and mild cognitive impairment," *The journals of gerontology. Series B, Psychological sciences and social sciences*, vol. 61, pp. P228–36, 08 2006.

[10] N. Dalia Cilia, T. D'Alessandro, C. De Stefano, and F. Fontanella, "Offline handwriting image analysis to predict alzheimer's disease via deep learning," in *2022 26th International Conference on Pattern Recognition (ICPR)*, 2022, pp. 2807–2813.

[11] N. D. Cilia, T. D'Alessandro, C. De Stefano, F. Fontanella, and M. Molinara, "From online handwriting to synthetic images for alzheimer's disease detection using a deep transfer learning approach," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 12, pp. 4243–4254, 2021.

[12] N. D. Cilia, T. D'Alessandro, C. De Stefano, and F. Fontanella, "Deep transfer learning algorithms applied to synthetic drawing images as a tool for supporting alzheimer's disease prediction," *Machine Vision and Applications*, vol. 33, no. 3, p. 49, 2022.

[13] N. D. Cilia, C. De Stefano, F. Fontanella, M. Molinara, and A. Scotto Di Freca, "Handwriting analysis to support alzheimer's disease diagnosis: A preliminary study," in *Computer Analysis of Images and Patterns*, M. Vento and G. Percannella, Eds. Cham: Springer International Publishing, 2019, pp. 143–151.

[14] N. D. Cilia, C. De Stefano, F. Fontanella, and A. Scotto di Freca, "Using genetic algorithms for the prediction of cognitive impairments," in *Lecture notes in computer science*, vol. 12104. Springer, 2020, pp. 479–493.

[15] G. Lozupone, E. Nardone, C. D. Pace, and T. D'Alessandro, "Transformers and cnns in neurodiagnostics: Handwriting analysis for alzheimer's diagnosis," in *International Conference on Pattern Recognition*. Springer, 2025, pp. 447–463.

[16] N. Cilia, C. De Stefano, F. Fontanella, and A. Scotto di Freca, "An experimental protocol to support cognitive impairment diagnosis by using handwriting analysis," *Procedia Computer Science*, vol. 141, pp. 466 – 471, 2018.

[17] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "'mini-mental state': A practical method for grading the cognitive state of patients for the clinician," *J. Psychiatric Res.*, vol. 12, no. 3, p. 189–198, 1975.

[18] A. Iavarone, B. Ronga, L. Pellegrino, E. Loré, S. Vitaliano, F. Galeone, and S. Carlomagno, "The frontal assessment battery (fab): Normative data from an italian sample and performances of patients with alzheimer's disease and frontotemporal dementia," *Funct Neurol.*, vol. 19, p. 191–195, 7 2004.

[19] Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, and H. Chertkow, "The montreal cognitive assessment, moca: A brief screening tool for mild cognitive impairment," *Journal of the American Geriatrics Society*, vol. 53, no. 4, p. 695–699, 4 2005.

[20] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[21] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

[22] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 11 976–11 986.

Alzheimer's Disease Screening with Limited and Noisy On-line Pentagon Drawings

Nina Hosseini-Kivanani
University of Luxembourg
Luxembourg
nina.hosseinikivanani@uni.lu

Elena Salobrar-García
Universidad Complutense de Madrid
Spain
elenasalobrar@med.ucm.es

Moises Diaz
Universidad de Las Palmas de Gran Canaria
Spain
moises.diaz@ulpgc.es

Miguel A. Ferrer
Universidad de Las Palmas de Gran Canaria
Spain
miguelangel.ferrer@ulpgc.es

Luis A. Leiva
University of Luxembourg
Luxembourg
luis.leiva@uni.lu

***Abstract*— Alzheimer's Disease (AD) is a progressive neurodegenerative disorder that leads to cognitive and motor decline. Handwriting analysis has gained attention as a non-invasive approach to detecting early neurodegenerative changes, leveraging cognitive and motor impairments associated with AD. This study explores on-line handwriting for AD screening using a standard neuropsychological assessment task (Pentagon drawings) where the collected data are both scarce and extremely noisy. We propose a novel data augmentation framework grounded in the Kinematic Theory to reconstruct noisy velocity profiles from scratch to generate realistic handwriting variations. We consider several classification tasks with Healthy Controls (HC), Mild Cognitive Impairment (MCI), and AD patients. Our results show that our proposed method achieves the best performance, offering thus a scalable, cost-effective, and non-invasive tool for early neurodegenerative disease screening, with practical applications in digital medicine and healthcare.**

Keywords*— *Alzheimer's disease, handwriting analysis, data augmentation, velocity profile reconstruction

I. Introduction

Neurodegenerative diseases progressively impair neural functioning, affecting cognition and motor control [1], [2]. Among them, Alzheimer's disease (AD) is the most prevalent and severe form of dementia, impacting millions worldwide [3]. While pharmacological treatments provide symptom relief, they do not stop disease progression. Therefore, early detection is critical for managing symptoms and slowing functional decline [4].

Existing diagnostic methods, including e.g. neuroimaging (MRI, PET) and cerebrospinal fluid (CSF) analysis, provide valuable clinical insights but remain costly, invasive, and inaccessible for routine screening [5], [6]. On the other hand, cognitive assessments offer a non-invasive alternative but are subject to variability due to anxiety, reduced patient cooperation, and subjective evaluation [7]. These limitations highlight the need for more objective and more accessible diagnostic tools.

Handwriting analysis has gained attention as a method for detecting early motor and cognitive impairments in AD [8]–[10], making it an interesting source of digital biomarkers [11], [12]. In this paper, we build upon this line of research and investigate Pentagon drawings, a standard neuropsychological assessment task. A particular challenge of our data is that they are both scarce and extremely noisy. We thus propose a novel data augmentation framework grounded in the Kinematic Theory to reconstruct the velocity profiles from scratch and generate realistic handwriting variations that we use for classification.

II. Related Work

Handwriting analysis has gained attention as a method for detecting early motor and cognitive impairments in AD [8]–[10], making it an interesting source of digital biomarkers [11], [12]. Several review and experimental studies have confirmed the relevance of this modality in AD screening, including a comprehensive review by De Stefano et al. [13], and more recent experimental work on spiral dynamics and transfer learning by Carfora et al. [14].

Some studies highlight that parameters such as writing speed, pen pressure, and movement fluency enable accurate classification of AD and mild cognitive impairment (MCI) patients [15], [16]. Computational models have used various features to differentiate AD patients from healthy individuals [17], [18]. Further, researchers have proposed Generative Adversarial Networks [19] and Diffusion models [20] for AD classification, however they require large datasets to be trained on, and most real-world datasets of AD patients are very limited in size. Recent work has investigated the Kinematic Theory with classic Machine Learning models such as Support Vector Machines [21]. To the best of our knowledge, we are the first to explore Deep Learning models in this context, since researchers have worked mostly with off-line data only [22]–[25].

The Kinematic Theory [26] and its latest instantiation, the Sigma-Lognormal model [27], describes handwriting movements by modeling velocity profiles using lognormal functions, and is among the most accurate approaches so far. Following this, we introduce velocity profile reconstruction and controlled distortion levels to create realistic samples, suitable AD screening. Our results show that our proposed method achieves the best performance, offering thus a scalable, cost-effective, and non-invasive tool for early neurodegenerative disease screening, with practical applications in digital medicine and healthcare.

III. Materials and Methodology

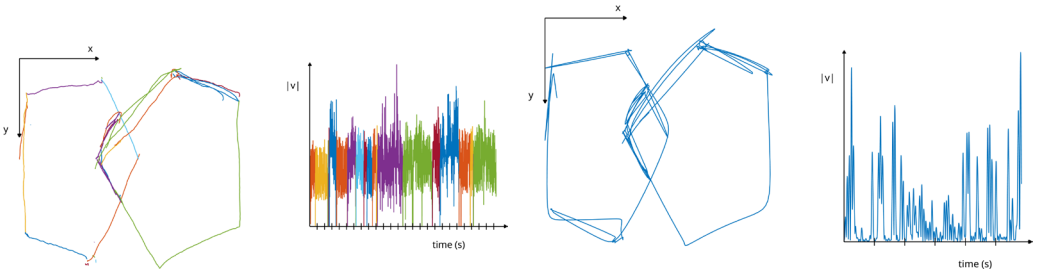
Previous work has used high-resolution digitizing tablets to capture data from AD patients, particularly through tasks like spiral drawing and handwriting analysis, which can reveal bradykinesia, hypometria, and sequencing impairments characteristic of the disease [16], [28]. However, these methods are considered unnatural to many users, who prefer the classic pen-and-pencil method [18]. To enhance ecological validity and user comfort, we used a 10.9-inch Repaper tablet[1], which enables the use of a real ballpoint pen on a blank sheet of paper while digitizing strokes in real time.

A standard pencil was instrumented with a magnetic ring that was sensed by the tablet, which sent the data via Bluetooth for later storage. This setup preserved a natural drawing experience while simultaneously recording on-line trajectories.

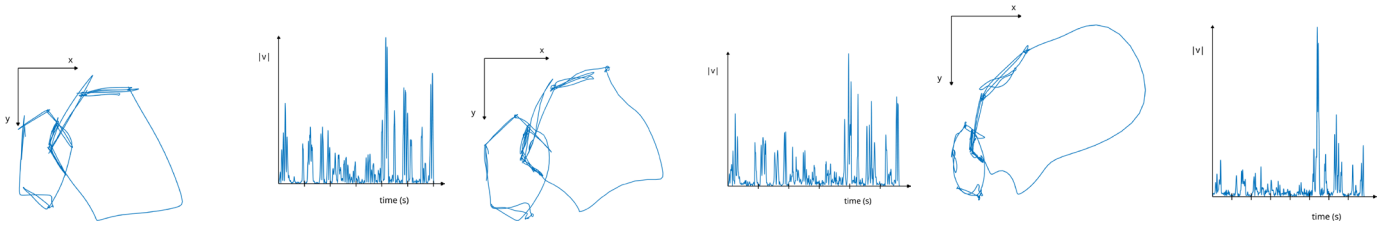
A. Data Collection

We collected 84 Pentagon drawings, commonly used in neuropsychological assessments [29], from 33 AD patients, 38 MCI patients, and 13 Healthy Controls (HCs). This corresponds to one drawing collected per participant.

The age of participants ranged between 76 and 83 years. The Repaper tablet provided the `x, y` coordinates without sampling rate information, so it was assumed to be 100 Hz. However, the velocity profiles were extremely noisy so they had to be entirely synthesized, as described next.



(a) Original trajectory (b) Reconstructed trajectory



(c) Synthetic trajectory, small distortion (d) Synthetic trajectory, medium distortion (e) Synthetic trajectory, large distortion

Fig. 1: Trajectories and velocity profiles of a Pentagon, originally drawn by an MCI patient.

B. Velocity Profile Reconstruction and Synthesis

Although the original trajectories contain multiple pen-down sequences (or *components*, see Fig. 1a), we used the entire trajectory without segmentation for reconstruction, as a single sequence, and ignoring the aforementioned sampling rate of 100 Hz. We applied iDeLog [30] for reconstruction, which is optimized for long and complex handwriting movements. First, iDeLog identifies virtual target points (high-curvature points) and designates them as stroke boundaries. Given that the timing of these points is known, iDeLog inserts a lognormal distribution between the start and end times of each stroke and resamples the trajectory. The lognormal distribution is centered at the midpoint of each stroke, ensuring that 98% of its area falls within the duration of the strokes, allowing for slight overlap between consecutive strokes. As a result, the reconstruction simulates a Pentagon drawing without lifting the pen from the paper, as shown in Fig. 1b.

Then, we synthesized new samples by applying a series of affine transformations [31] to the reconstructed data, in the following order:

1) Gaussian noise distortion, centered on each point with variance in $(1, 5, 10) \times 10^{-3}$, according to three chosen distortion levels; i.e., small, medium, and large.
2) Sinusoidal function [32] to modify stroke lengths non-uniformly. The sinusoid periods are selected from the intervals $[(0.6, 1.2), (0.5, 1.3), (0.6, 1.4)]$, while the corresponding amplitude ranges are $[(11, 14), (8, 17), (5, 20)]$, each representing one of the three distortion levels.

[1] https://www.iskn.co/eu

3) Perspective distortion, with a shearing effect with values from $[(4,5),(2,7),(0,9)]\times10^{-2}$ and depth-related size changes adjusted within the range $[(5,7),(2,9.9),(-1,13)]\times10^{-4}$.
4) Rotational distortion, applied randomly in either a positive or negative direction. Positive rotation angles are sampled from $[(5,20),(10,30),(15,40)]$ deg, while negative rotations are drawn from $[(-20,-5),(-30,-10),(-40,-15)]$ deg.
5) Finally, the trajectory is either enlarged or shrunk based on randomly chosen scaling factors. The scaling intervals are $[(7,9),(5,1),(3,1.1)]\times10^{-1}$ or $[(1.1,1.3),(0.9,1.5),(0.7,1.7)]$, depending on the transformation type.

A normally distributed random value was sampled for each parameter within the specified intervals. Then, the resulting trajectory was then parameterized using iDeLog. Subsequently, the virtual target point positions were modified. The minimum distance m between adjacent virtual target points (vtp) is calculated. A circular region with radius $m\cdot\delta_r$ is generated, where $\delta_r\in(0.1,0.3,0.5)$ varies according to a distortion level (small, medium, large). The new virtual target point positions (vtp') are then computed as follows:

$$vtp'_x = vtp_x + m\cdot\delta_r\cdot r\cdot\cos(2\pi r) \qquad vtp'_y = vtp_y + m\cdot\delta_r\cdot r\cdot\sin(2\pi r) \tag{1}$$

where r is a randomly generated factor within $(0,1)$, introducing thus a controlled distortion.

The percentage of stroke duration was then randomly modified within the intervals $\delta_t\in\pm(10\%,30\%,50\%)$, affecting the total duration. The same percentage modification is applied to the σ_i parameter of each lognormal component in the Pentagon decomposition. Due to the relationship between μ and σ in neuromuscular systems, the μ_i values were updated as:

$$\mu'_i = \ln(\exp(\mu_i - \sigma_i^2)) + \sigma'_i \tag{2}$$

where $\sigma'_i = \sigma_i\cdot\delta_t$. Using these updated parameters, the final trajectory was synthesized. Figures 1c to 1e provide visual examples of synthesized data with small, medium, and large distortions for a randomly selected Pentagon drawing.

IV. Experiments

We used the same classifier in all experiments, informed by previous work [33]: a GRU model with a single layer of 200 units (embedding size) using tanh activation and 0.15 dropout. The model takes as input the offsets between consecutive points $(\Delta x,\Delta y)$ to make the trajectories scale-invariant.

The model was trained on half of the data (stratified sampling) for 100 epochs with early stopping of 10 epochs, using Adam optimizer (learning rate $\eta=0.001$ and decay rates $\beta_1=0.9$, $\beta_2=0.999$). Only the training data was augmented, leaving the test data as an unseen partition for final model evaluation.

We compared our method against no data augmentation and a jitter-based method using Gaussian noise centered at each coordinate $\mathcal{N}(\mu=0,\sigma=5)$, as described in previous work [18], [34]. The value of $\sigma=5$ was set according to the mean value of the offsets between consecutive points observed in our dataset. We generated 50 variations per human sample with each method. The results of our experiments are summarized in Table I.

TABLE I: Performance metrics on the test data. Best results highlighted in bold.

Method	Adj. Precision (%)	Adj. Recall (%)	Adj. F-measure (%)	Accuracy (%)	AUC ROC (%)
Classification Task: Healthy Controls vs. AD patients					
No augmentation	48.39	69.57	57.08	69.57	50.00
Jitter baseline	55.21	60.87	57.21	60.87	47.77
Ours, small distortions	88.32	88.46	88.23	88.46	84.72
Ours, medium distortions	**96.58**	**96.15**	**96.21**	**96.15**	**97.22**
Ours, large distortions	74.11	73.08	73.48	73.08	70.14
Classification Task: Healthy Controls vs. MCI patients					
No augmentation	82.46	76.92	69.84	76.92	57.14
Jitter baseline	82.46	76.92	69.84	76.92	57.14
Ours, small distortions	86.44	84.62	85.02	84.62	85.42
Ours, medium distortions	**97.37**	**97.22**	**97.22**	**97.22**	**97.22**
Ours, large distortions	65.73	69.23	65.64	69.23	56.94
Classification Task: MCI vs. AD patients					
No augmentation	64.47	63.89	62.88	63.89	63.00
Jitter baseline	63.82	63.89	63.80	63.89	63.62
Ours, small distortions	83.75	83.33	83.28	83.33	83.33
Ours, medium distortions	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Ours, large distortions	59.03	58.33	57.51	58.33	58.33

V. Discussion and Future Work

Our experiments show that it is possible to identify AD and MCI patients from HCs with outstanding performance using scarce and noisy on-line data, outperforming previous work on similar tasks [18], [34], [35] as well as previous work that used off-line data [22]–[25]. We have found that medium-level distortions work best.

Although we did not perform an exhaustive grid search over the augmentation parameter space, we systematically evaluated three representative distortion levels—small, medium, and large—which were sufficient to identify the optimal performance range across all classification tasks. These levels were selected to ensure a balance between meaningful variability and physiological plausibility, and they yielded consistent trends throughout our experiments.

With respect to the sinusoidal function used to modulate stroke lengths, we acknowledge that its period values were fixed in this study. Nevertheless, previous work [36], [37] has shown that variations in the sinusoidal period exert minimal influence on the resulting handwriting synthesis and its downstream classification utility. Therefore, our selection of period intervals is consistent with these prior findings and justified within the scope of this study.

The crux of our method lies in working on the sigma-lognormal space. That is, working with reconstructions of human samples and synthetically generated samples from those reconstructions. Future work will explore other drawing symbols used in neuropsychological assessment tasks, such as Clocks and Houses, as well as multi-class classification tasks. Further, our GRU model is very lightweight, so we plan to deploy it in clinical settings as part of a mobile app.

VI. Conclusion

We have explored on-line handwriting for AD screening using a standard neuropsychological assessment task (Pentagon drawings) where the collected data are both scarce and extremely noisy. We have proposed a novel data augmentation framework grounded in the Kinematic Theory to reconstruct noisy velocity profiles from scratch to generate realistic handwriting variations. Through several classification tasks, our results show that our proposed method achieves the best performance.

While promising, our approach has certain limitations. Notably, the augmentation parameters were fixed and not optimized for generalizability. Thus, further sensitivity analysis and validation are needed to ensure robustness across different settings and symbols. In this light, our method can be viewed as a potentially scalable, cost-effective, and non-invasive aid for early neurodegenerative disease screening, with future applications in digital medicine and healthcare pending further validation.

Acknowledgments

Research supported by PID2023-146620OB-I00 (funded by MICIU/AEI 10.13039/501100011033), the European Union's FEDER program, CajaCanaria and LaCaixa (grant 2023DIG05), and the European Innovation Council under grant 101071147.

References

[1] C. P. Fernandes, G. Montalvo, M. Caligiuri, M. Pertsinakis, and J. Guimaraes, "Handwriting changes in Alzheimer's disease: A systematic review," *J. Alzheimer's Dis.*, vol. 96, no. 1, 2023.
[2] N. D. Cilia, T. D'Alessandro, C. De Stefano, F. Fontanella, and M. Molinara, "From online handwriting to synthetic images for Alzheimer's disease detection using a deep transfer learning approach," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 12, 2021.
[3] H. Qi, R. Zhang, Z. Wei, C. Zhang, L. Wang, Q. Lang, K. Zhang, and X. Tian, "A study of auxiliary screening for Alzheimer's disease based on handwriting characteristics," *Front. Aging Neurosci.*, vol. 15, 2023.
[4] C. A. Anacur, A. G. Yilmaz, and M. Aykut, "Handwriting analysis for Alzheimer's diagnosis: Feature enrichment and machine learning classification," in *Proc. IDAP*, 2024.
[5] S. Shimizu, D. Hirose, H. Hatanaka, N. Takenoshita, Y. Kaneko, Y. Ogawa, H. Sakurai, and H. Hanyu, "Role of neuroimaging as a biomarker for neurodegenerative diseases," *Front. Neurol.*, vol. 9, 2018.
[6] S. M. McGinnis, "Neuroimaging in neurodegenerative dementias," *Seminars in neurology*, vol. 32, no. 4, 2012.
[7] H. Hampel, L. M. Shaw, P. Aisen, C. Chen, A. Lleó, T. Iwatsubo, A. Iwata, M. Yamada, T. Ikeuchi, J. Jia *et al.*, "State-of-the-art of lumbar puncture and its place in the journey of patients with Alzheimer's disease," *Alzheimer's & Dementia*, vol. 18, no. 1, 2022.
[8] S. Toffoli, F. Lunardini, C. G. de Isla, and S. Ferrante, "AI-based ecological monitoring of handwriting to early detect cognitive decline," in *Proc. BHI*, 2023.
[9] U. Mitra and S. U. Rehman, "ML-powered handwriting analysis for early detection of Alzheimer's disease," *IEEE Access*, 2024.
[10] Y. Luo and Z. Yang, "DynGAN: Solving mode collapse in GANs with dynamic clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2024.
[11] M. Vimaladevi, R. Thangamani, S. Priyadharshini, B. Varshini, and A. Tarun, "Prediction of Alzheimer's disease by analyzing handwriting dynamics using machine learning algorithms," in *Proc. ICESC*, 2024.
[12] N. D. Cilia, C. De Stefano, F. Fontanella, M. Molinara, and A. Scotto Di Freca, "Handwriting analysis to support Alzheimer's disease diagnosis: a preliminary study," in *Proc. CAIP*, 2019.
[13] C. De Stefano, F. Fontanella, D. Impedovo, G. Pirlo, and A. S. di Freca, "Handwriting analysis to support neurodegenerative diseases diagnosis: A review," *Pattern Recognit. Lett.*, vol. 121, 2019.
[14] D. Carfora, S. Kim, N. Houmani, S. Garcia-Salicetti, and A.-S. Rigaud, "On extracting digitized spiral dynamics' representations: A study on transfer learning for early alzheimer's detection," *Bioengineering*, vol. 9, no. 8, p. 375, 2022.
[15] P. Ghaderyan, A. Abbasi, and S. Saber, "A new algorithm for kinematic analysis of handwriting data; towards a reliable handwriting-based tool for early detection of Alzheimer's disease," *Expert Syst. Appl.*, vol. 114, 2018.
[16] C. Kahindo, M. El-Yacoubi, S. Garcia-Salicetti, V. Cristancho-Lacroix, H. Kerhervé, and A.-S. Rigaud, "Semi-global parameterization of online handwriting features for characterizing early-stage Alzheimer and mild cognitive impairment," *IRBM*, vol. 39, no. 6, 2018.
[17] J. A. Meltzer, "Towards early prediction of Alzheimer's disease through language samples," *EClinicalMedicine*, vol. 29, 2020.
[18] N. Hosseini-Kivanani, E. Salobrar-García, L. Elvira-Hurtado, M. Salas, C. Schommer, and L. A. Leiva, "Blueprint of tomorrow: Contrasting off-line and on-line drawing tasks for Alzheimer's disease screening," in *Proc. IDEAL*, 2024.

[19] B. Wang, J. Lan, and J. Gao, "Multi-class GAN for generating multi-class images in object recognition," *J. Opt. Soc. Am. A*, vol. 39, no. 5, 2022.
[20] X. Yang, T. Ye, X. Yuan, W. Zhu, X. Mei, and F. Zhou, "A novel data augmentation method based on denoising diffusion probabilistic model for fault diagnosis under imbalanced data," *IEEE Trans. Ind. Inform.*, 2024.
[21] T. D'Alessandro, C. Carmona-Duarte, C. De Stefano, M. Diaz, M. A. Ferrer, and F. Fontanella, "A machine learning approach to analyze the effects of Alzheimer's disease on handwriting through lognormal features," in *Proc. IGC*, 2023.
[22] S. Chen, D. Stromer, H. A. Alabdalrahim, S. Schwab, M. Weih, and A. Maier, "Automatic dementia screening and scoring by applying deep learning on clock-drawing tests," *Sci. Rep.*, vol. 10, no. 1, 2020.
[23] N. Ruengchaijatuporn, I. Chatnuntawech, S. Teerapittayanon, S. Sriswasdi, S. Itthipuripat, S. Hemrungrojn, P. Bunyabukkana, A. Petchlorlian, S. Chunamchai, T. Chotibut, and C. Chunharas, "An explainable self-attention deep neural network for detecting mild cognitive impairment using multi-input digital drawing tasks," *Alzheimers Res. Ther.*, vol. 78, no. 14, 2022.
[24] R. Raksasat, S. Teerapittayanon, S. Itthipuripat, K. Praditpornsilpa, A. Petchlorlian, T. Chotibut, C. Chunharas, and I. Chatnuntawech, "Attentive pairwise interaction network for ai-assisted clock drawing test assessment of early visuospatial deficits," *Sci. Rep.*, vol. 13, no. 1, 2023.
[25] N. Hosseini-Kivanani, I. C. Oliveira, S. Kilinç, and L. A. Leiva, "Efficient automatic data augmentation of CDT images to support cognitive screening," in *Proc. ICAART*, 2025.
[26] R. Plamondon, "A kinematic theory of rapid human movements: Part ii. movement time and control," *Biol. Cybern.*, vol. 72, 1995.
[27] R. Plamondon and M. Djioua, "A multi-level representation paradigm for handwriting stroke generation," *Hum. Mov. Sci.*, vol. 25, no. 4, 2006.
[28] M. Kobayashi, Y. Yamada, K. Shinkawa, M. Nemoto, K. Nemoto, and T. Arai, "Automated early detection of alzheimer's disease by capturing impairments in multiple cognitive domains with multiple drawing tasks," *Journal of Alzheimer's Disease*, vol. 88, no. 3, pp. 1075–1089, 2022.
[29] E. Helmes, "Cognitive screening of older adults: the utility of pentagon drawing," *International psychogeriatrics*, vol. 25, no. 3, 2013.
[30] M. A. Ferrer, M. Diaz, C. Carmona-Duarte, and R. Plamondon, "iDeLog: iterative dual spatial and kinematic extraction of sigma-lognormal parameters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, 2018.
[31] M. Maslych, E. M. Taranta, M. Aldilati, and J. J. Laviola, "Effective 2d stroke-based gesture augmentation for RNNs," in *Proc. CHI*, 2023.
[32] M. Diaz, A. Fischer, R. Plamondon, and M. A. Ferrer, "Towards an automatic on-line signature verifier using only one reference per signer," in *Proc. ICDAR*, 2015.
[33] L. A. Leiva, M. Diaz, M. A. Ferrer, and R. Plamondon, "Human or machine? it is not what you write, but how you write it," in *Proc. ICPR*, 2020.
[34] N. Hosseini-Kivanani, E. Salobrar-García, L. Elvira-Hurtado, M. Salas, C. Schommer, and L. A. Leiva, "Predicting Alzheimer's disease and mild cognitive impairment with off-line and on-line house drawing tests," in *Proc. eScience*, 2024.
[35] N. Hosseini-Kivanani, E. Salobrar-García, L. Elvira-Hurtado, I. López-Cuenca, R. de Hoz, J. M. Ramírez, P. Gil, M. Salas, C. Schommer, and L. A. Leiva, "Better together: Combining different handwriting input sources improves dementia screening," in *Proc. eScience*, 2023.
[36] M. A. Ferrer, M. Diaz, C. Carmona-Duarte, and A. Morales, "A behavioral handwriting model for static and dynamic signature synthesis," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1041–1053, 2016.
[37] M. Diaz, M. A. Ferrer, G. S. Eskander, and R. Sabourin, "Generation of duplicated off-line signature images for verification systems," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 5, pp. 951–964, 2016.





Special Session

KNM2 & More for Potential Collaborations

Lognormal Function as a Primitive for Sinusoidal Waveforms: Case of the Velocity Profile in Oscillatory Human Movement

Moussa Djioua

Département de Génie Électrique,Polytechnique Montréal
2500 Chemin de Polytechnique, Montréal,H3T 1J4,Québec, Canada
m.djioua@polymtl.ca

***Abstract*— This paper summarizes part of a broader study on oscillatory human movement, considered a special case of composite movement where successive stroke directions are in perfect opposition, and their amplitudes are maintained at a constant magnitude. The representation of oscillatory movement using the Sigma-Lognormal model is examined. Over the past decades, the velocity profile of human movement has been extensively studied and modeled as a biological signal. It has been demonstrated that its bell-shaped profile converges to the lognormal function, which serves as a fundamental primitive from which all other causal signals are constructed. Specifically, it is shown that oscillations emerging in biological systems originate from the manifestations of the central limit theorem and agonist-antagonist synergy. This study begins by demonstrating that the purest form of non-causal oscillation, the sinusoid, results from the infinite concatenation of Gaussian functions. To accurately represent causal biological oscillations, the sinusoid is effectively approximated by the infinite concatenation of lognormal functions. The analysis results highlight that, to approximate sinusoidal oscillations, the human motor system regulates agonist-antagonist synergy under the minimum time principle. This finding enhances the understanding of how oscillations are utilized to represent composite movements such as handwriting and signature.**

Keywords*— *Oscillatory human movement,Sigma-Lognormal model,Kinematic Theory, Minimum-Time principle,

I. Introduction

Since the observation of the stereotypical tangential velocity profile in rapid human movement, several models have been proposed to represent its bell-shaped form, which is generally asymmetric [1], [2]. Either directly or indirectly, all these models rely on probability density functions, derived from minimization principles such as minimum time, jerk, or energy. Using the Central Limit Theorem, the proportionality hypothesis, and agonist-antagonist synergy, Plamondon et al. [3]–[5] provided mathematical evidence and empirical validation—through electromyography (EMG) signals—demonstrating that the lognormal profile is the most suitable causal bell-shaped function to describe the velocity profile of rapid movements. This led to the development of the Sigma-Lognormal model, which was designed to represent composite movements such as handwriting and signature production [9], [10]. Furthermore, the pseudo-oscillatory nature of planar handwriting and signature trajectories has inspired models based on coupled oscillators to generate these movement patterns [6], [7]. The study of simple pointing movements has been approached through a detailed analysis of rapid movements, as their tangential velocity profile is unimodal, and the influence of movement direction on the velocity shape is negligible. Based on this, research has focused on identifying links between different models proposed to describe the bell-shaped velocity profile. It has been shown that the lognormal profile represents the asymptotic limit toward which all other models converge [8]. Before investigating handwriting and signature movements, the same approach was applied to oscillatory movements at maximum speed, aiming to obtain a velocity profile approximating a sinusoidal waveform. By considering oscillatory movement as a causal biological signal, disregarding directional effects, and assuming that two consecutive strokes are perfectly opposing and curvature-free, it was demonstrated that the Sigma-Lognormal model effectively captures the velocity profile of oscillatory movements.

II. Oscillatory movement

Handwriting and signature movements can be interpreted as being generated by a sinusoidal oscillator, whose amplitude, frequency, and phase are modulated to produce the desired trajectory [6], [7]. The coordinates of these trajectories correspond to the projections of the oscillator onto the Euclidean axes. However, the oscillatory nature of the movement is not inherently contained in the projection itself but rather in the projected velocity. The structural characteristics of handwritten symbols can be shaped by appropriately modulating the curvature of the trajectory produced by the oscillatory movement of the end-effector. An oscillatory movement consists of rapid back-and-forth displacements between two positions. By controlling the initial direction, curvature, and amplitude of each rapid movement, the system can generate desired stroke patterns. Due to its agonist-antagonist structure, the peripheral motor system is considered to operate in an oscillatory mode, driven by a control system that modulates amplitude, phase, and frequency to produce the intended trajectory. In the absence of active control, the Cartesian coordinate profiles of the generated trajectory would exhibit regular oscillations with relatively constant amplitudes, close to a sinusoidal vibration.

In the study of simple motor actions, ballistic movement represents a special case that has provided insights into the intrinsic behavior of both control and peripheral components of the motor system. Following this reasoning, oscillatory movement at maximum speed can be considered a special case of composite movement, allowing for the characterization of intrinsic properties of the motor system's components.

A sinusoidal vibration is a perpetual oscillatory signal with infinite energy. It is non-causal and possesses mathematical properties that have made the Fourier transform a powerful tool for signal analysis and processing, offering an alternative representation of the information contained within the signal. Any signal can be decomposed into an infinite sum of sinusoidal oscillations. While, from a mathematical perspective, this decomposition provides an unparalleled analytical framework for frequency-domain representation, from a physical and biological standpoint, constructing a finite-energy signal by summing an infinite number of infinite-energy signals is inherently counterproductive. Moreover, this approach raises questions regarding the appropriateness of representing a causal signal using non-causal components.

Instead of representing a finite-energy signal as an infinite sum of infinite-energy signals, we propose reversing the approach: considering the sinusoidal waveform (which has infinite energy) as the result of an infinite concatenation—rather than superposition—of a single, finite-energy primitive.

A. Case of non-causal primitives

First, we consider that the Gaussian function is used to fit a temporal signal. It serves as the atomic primitive for non-causal signals, characterized by an arithmetic mean τ and a standard deviation σ such that

$$\mathcal{N}(t,\tau,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{t-\tau}{\sigma\sqrt{2}}\right)^2} \tag{1}$$

Let $\cos(2\pi f t)$ be a sinusoidal vibration of frequency f, and let $\mathrm{gcos}(t)$ be the concatenation of an infinite number of Gaussian functions, such that:

$$\mathrm{gcos}(t,\tau,\sigma) = K \sum_{n=-\infty}^{+\infty} (-1)^n \mathcal{N}(t, n\tau, \sigma) \tag{2}$$

With K as a normalization coefficient. In a sinusoidal vibration of frequency f, the interval between the maximum and the minimum is equal to $1/2f$. To represent the cosine function using the $gcos$ function, the condition $\tau = \frac{1}{2f}$ must be satisfied. To obtain a complete oscillation period, a concatenation of three Gaussian functions is required. Given that the total duration of a single Gaussian is fixed at 6σ, we define $\sigma = \frac{1}{2f}$. This leads to a representation of sinusoidal oscillation as an infinite concatenation of shifted Gaussian functions, with parameter $x = 2\pi f t$, whose expression is given by (see Fig.1.a-b):

$$\cos(x) \sim \frac{1}{\vartheta_4\left(0, e^{-\frac{1}{2}}\right)} \sum_{n=-\infty}^{+\infty} (-1)^n e^{-\frac{1}{2\pi^2}(x-n\pi)^2} \tag{3}$$

$$\sin(x) \sim \frac{1}{\vartheta_4\left(0, e^{-\frac{1}{2}}\right)} \sum_{n=-\infty}^{+\infty} (-1)^n e^{-\frac{1}{2\pi^2}\left(x-(n+\frac{1}{2})\pi\right)^2} \tag{4}$$

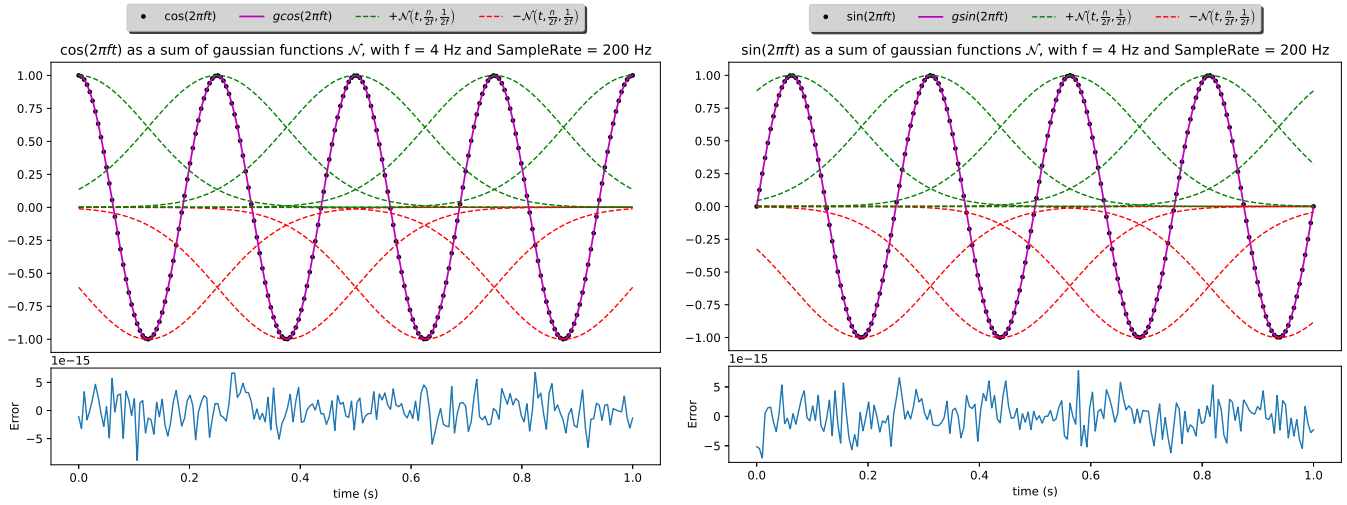
With, ϑ_4, the Jacobi Theta function of order 4, with $u = 0$ and $q = e^{-\frac{1}{2}}$, given by:

$$\vartheta_4(u,q) = 1 + 2\sum_{n=1}^{+\infty} (-1)^n q^{n^2} \cos(nu) \tag{5}$$

B. Discussion

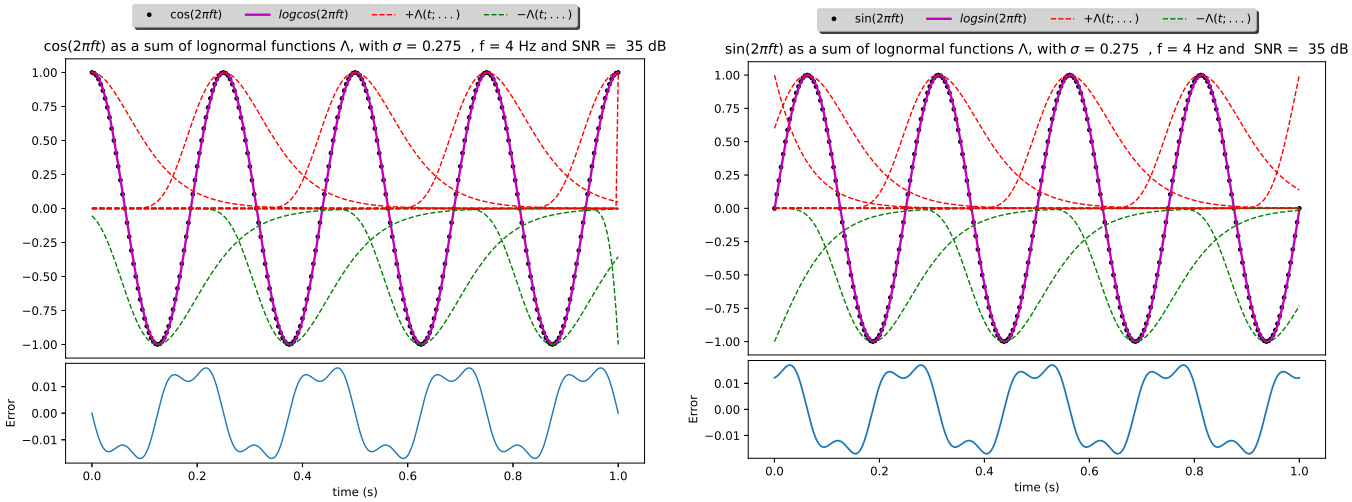
The representation of a sinusoidal oscillation as an infinite sum of Gaussians leads to the following observations:

- As illustrated in Fig. 1-a, the fitting error between the function $gcos$ and the cosine function is on the order of 10^{-14}. This error is independent of both the sample rate and the oscillation frequency and is considered to arise from numerical precision limitations. The same accuracy is observed in the power series expansion of $gcos$, where the even-order coefficients match those of the cosine function's expansion, while the odd-order coefficients tend toward zero.
- In the implementation of the infinite sum of Gaussians, the theoretical infinity (∞) is approximated by the finite value 8. Beyond this limit, the noise-like shape and the amplitude of the error remain unchanged. If this limit is further reduced, a deterministic error emerges, taking the form of an oscillation with a significantly increasing amplitude at the boundaries. For example, if the sinusoidal oscillation has a frequency of 4 Hz over a time window W_t of one second, then $1 + 2|fW_t| = 9$ Gaussian functions are required. Additionally, 8 Gaussians must be added at the beginning and end



(a) (b)

Fig. 1. Representations of a sinusoidal oscillation by the concatenation of Gaussian functions for (a) cosinus and (b) sinus functions



(a) (b)

Fig. 2. Representations of a sinusoidal oscillation by the concatenation of Lognormal functions for (a) cosinus and (b) sinus functions

of the signal to account for the influence of those outside the time window. Consequently, a total of 25 Gaussians are needed to simulate a trigonometric oscillation lasting one second. In the general case, the number of Gaussians required to approximate a sinusoidal signal of frequency f and duration W_t is given by: $N = 17 + 2|fW_t|$.

- Given that the value of $\sigma = \frac{1}{2f}$, the duration of each Gaussian function, which equals 6σ, spans three periods of the sinusoidal vibration. Additionally, since $\tau = \sigma = \frac{1}{2f}$, the shift τ of each Gaussian component, occurring at 2σ from the origin, aligns precisely with the first inflection point of the preceding component. Thus, when observing a vibration of any frequency f, it should be understood as the result of an underlying synergistic interaction between two agonist and antagonist systems, where the response duration of each is three times slower than that of the vibration. The agonist-antagonist coupling operates as follows: the agonist activates its antagonist upon reaching its inflection point, corresponding to the first peak of its derivative, and vice versa. Since the vibration represents a velocity profile, the antagonist is triggered when the agonist reaches its maximum acceleration.

III. Case of causal primitives

However, the velocity of an oscillatory movement is a causal biological signal, and the lognormal function can serve as its primitive, as be done for velocity of rapid movement. Given that the lognormal function is inherently causal and asymmetric, it can only provide an approximate representation of a symmetric sinusoidal oscillation. Finding an appropriate approximation requires defining initial constraints. If the sole criterion were the minimization of the fitting error between the cosine function and its lognormal-based approximation, the solution would converge toward $\sigma \to 0$, where the lognormal function asymptotically approaches a Gaussian. Instead, the chosen constraint is the application of a minimization principle based on a physical quantity, such as time, jerk, or energy.

The lognormale function $\Lambda(t, t_0, \tau, \sigma)$ is given by :

$$\Lambda(t - t_0, \tau, \sigma) = \frac{1}{\sigma(t - t_0)\sqrt{2\pi}} e^{-\left(\log\left(\frac{t-t_0}{\tau}\right)/\sigma\sqrt{2}\right)^2} \tag{6}$$

Where τ is a geometric mean, or median, σ the asymmetry coefficient,t_0 represents the occurrence or the shift of the function along the time axis. These are parameters inherited from the Gaussian function. Let $logcos$ be the approximation of cosinus and $logsin$ be the approximation of sinus functions, constructed by the infinite concatenation of lognormal functions, such that (see Fig.2.a-b):

$$\text{logcos}(t, f, \tau, \sigma) = K \sum_{n=-\infty}^{\infty} (-1)^n \Lambda(t - t_n, \tau, \sigma) \tag{7}$$

$$\text{logsin}(t, f, \tau, \sigma) = K \sum_{n=-\infty}^{\infty} (-1)^n \Lambda\left(t - t_n + \frac{3}{4f}, \tau, \sigma\right) \tag{8}$$

With K as a normalization coefficient and, t_n a time shift such that:

$$t_n = \frac{n}{2f} - \tau e^{-\sigma^2} \tag{9}$$

We aim to determine the combinations of the two parameters (τ, σ) that minimize the fitting error between $logcos$ and cosinus functions. This error quantifies the approximation accuracy and is measured in decibels (dB) using the signal-to-noise ratio (SNR), defined as:

$$SNR(\tau, \sigma) = 10 \log_{10} \left[\frac{\int_t \cos^2(2\pi f t) dt}{\int_t \left[\cos(2\pi f t) - \text{logcos}(t, f, \tau, \sigma)\right]^2 dt} \right] \tag{10}$$

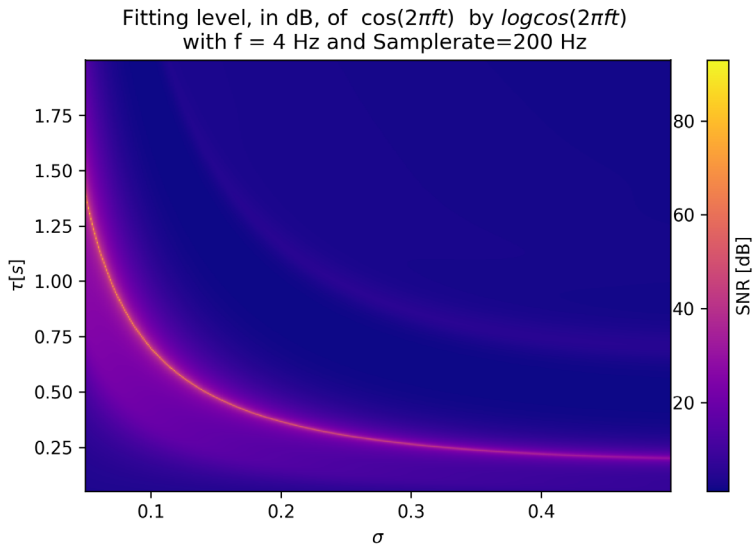
In Fig. 3-a, some combinations of the two parameters form a characteristic ridge, corresponding to an SNR approximately greater than or equal to 25 dB. However, there exist infinitely many possibilities to reach an approximation with a higher SNR value. The combinations of (τ, σ) that lead to local SNR maxima result in a U-shaped relationship between the duration $T = \tau e^{3\sigma}$ and σ , as shown in Fig. 3-b. Regardless of the oscillation frequency, adjusting the duration T as a function of σ for different frequency values (ranging from 1.2 to 16 Hz) yields an optimal value of σ that is approximately $\sigma_{opt} = 0.275$. This corresponds to the minimum-time principle, a concept frequently observed in the behavior of biological systems.

This implies that, for this optimal value , the relationship between the parameter τ and the frequency f is given by: $\tau \approx \frac{1.1276}{f}$.

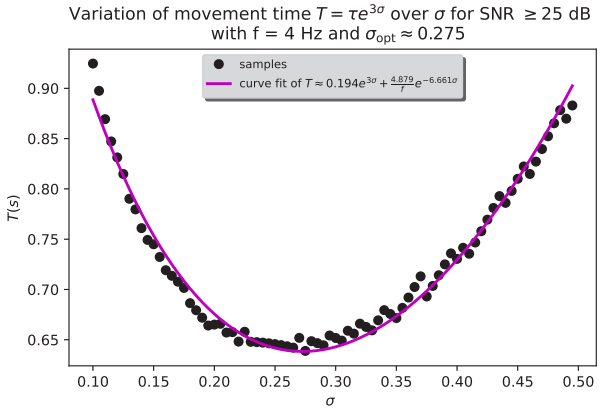
Let us consider the case where the activation of the antagonist at time t_{n+1} coincides with the reaction time of the agonist at $t_n + \tau e^{-3\sigma}$, such that $t_{n+1} = (t_n + \tau e^{-3\sigma}) \to \tau e^{-3\sigma} = \frac{1}{2f} \to \sigma \approx 0.27$. This results in a value close to σ_{opt}, associated with oscillations generated in minimum-time, where the antagonist is automatically triggered by the agonist as soon as the latter begins to respond mechanically, i.e, at a reaction time.

IV. Conclusion

This study shows that, in the case of non-causal signals, oscillation arises from the synergy of two systems in perfect opposition, where the free or impulse response is a Gaussian function. One system is activated when the acceleration of the other reaches its maximum. However, in the case of causal biological systems, oscillation emerges from a similar type of synergy, where the free response of the systems is a lognormal profile. Here, one system is triggered at the moment corresponding to the so-called mechanical reaction time of the other. This study employed mathematical and simulation tools to model the velocity profile of an oscillatory movement, independent of its spatial information, using the Sigma-Lognormal function. The results indicate that to achieve a more accurate approximation of its ideal form, the sinusoid, the motor control system operates according to the minimum-time principle. Furthermore, these findings highlight the importance of considering the hidden concatenation of lognormal primitives when fitting real movement data. Additionally, as previously established for simple movements, this study aims to establish connections between the various models proposed for representing handwriting



(a) Variations of the fitting error as a function of τ and σ



(b) Variations of a duration $T = \tau e^{3\sigma}$ as a function of σ

Fig. 3. Search for the parameters τ and σ that provide an approximation of pure oscillations with $SNR \geq 25$ dB.

and signature composite movements with concatenation and oscillation paradigms. From the perspective of causality and the finite energy of biological signals, we have demonstrated that the basis used to decompose these signals, such Fourier transforms, wavelets, etc. are themselves compositions of the gaussian and lognormal primitives functions, which emerge from the transition of stochastic functioning to deterministic behavior, as described by the central limit theorem. This suggests that the lognormal function can serve as a baseline primitive for analyzing and decomposing causal signals into fundamental components.

References

[1] Abend, W., Bizzi, E. and Morasso, P. "Human arm trajectory formation", Brain 105, 331–348, 1982.
[2] Morasso, P. "Spatial control of arm movements", Experimental Brain Research 42, 223–227 1981.
[3] Plamondon,R."A kinematic theory of rapid human movements:PartI:Movement representation and generation". Biological Cybernetics..72:295-307,1995.
[4] Plamondon,R."A kinematic theory of rapid human movements:PartII.Movement time and control". Biological Cybernetics. 72:309-320,1995.
[5] Plamondon,R.et al."A kinematic theory of rapid human movement.Part IV:a formal mathematical proof and new insights".Biol.Cybern. 89,126–138 2003.
[6] Teulings, H.L., Thomassen, A.J.W.M., Schomaker, L.R.B., Morasso, P.: Experimental protocol for cursive script acquisition: the use of motor information for the automatic recognition of cursive script. In: Report 3.1.2., ESPRIT Project, 419 1986.
[7] Morasso,P., Casadio, M., Mohan, V.,Sanguineti, V. Self-organizing topological maps and motor primitives in the brain. 2015.
[8] Djioua, M. Plamondon R "The limit profile of a rapid movement velocity", Human Movement Science 2009.
[9] Djioua, M. O'Reilly, C. and Plamondon, R."An interactive trajectory synthesizer to study outlier patterns in handwriting recognition and signature verification," Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), vol. 1, pp. 1124- 1127. 2006.
[10] O'Reilly, C. , Plamondon, R. "Development of a sigma-lognormal representation for on-line signatures". Pattern Recognit Spec Issue Front Handwrit Recognit. Pattern Recognition. 42. 3324-3337. 10.1016 .2008.

The Pen Stroke Test to Identify Suboptimal Neurodevelopment in Children Born Prematurely

Thuy M. Luu[1,2], Guillaume Séguin de Broin[3], Mi-Suk Kang-Dufour[1], Victoria Côté-Piché[1], Léanne Brabant[1], Marie-Michèle Gagnon[1], Mélanie Gagnon[1], Catherine Bernard[1], Véronique Belval[1], Marie-Noëlle Simard[1,4,*], Réjean Plamondon[3,*], and Mathieu Dehaes[1,5,6,*]

[1]Research Centre, Sainte-Justine University Hospital Centre, Montreal, Canada. [2]Department of Pediatrics, Sainte-Justine University Hospital Centre, Université de Montréal, Montreal, Canada. [3]Department of Electrical Engineering, Polytechnique Montreal, Montreal, Canada. [4]School of Rehabilitation, Université de Montréal, Canada. [5]Institute of Biomedical Engineering, Université de Montréal, Montreal, Canada.

[6]Department of Radiology, Radio-oncology and Nuclear Medicine, Université de Montréal, Montreal, Canada.

[*]Equal contribution as co-senior. Corresponding authors: rejean.plamondon@polymtl.ca & mathieu.dehaes@umontreal.ca

***Abstract*—**The development of screening strategies to identify premature children with potential neurodevelopmental problems is necessary to allow early clinical intervention and access to specialized services. The objective of this study was to assess the Pen Stroke Test on digitizer (PST) as a developmental screening tool in preschool-aged children born prematurely. 75 infants born at 29-36 weeks' gestational age prospectively underwent the PST at age 4.5 years to assess motor development. Derived metrics included the signal-to-noise ratio (SNR) and the number of lognormal impulses (NbLog) measuring the quality and the fluidity of the child pen stroke, respectively. Primary outcome was neurodevelopment at corrected age 4.5 years. SNR was positively associated with manual dexterity and motor skills while NbLog was negatively associated with manual dexterity, which is aligned with the kinematic theory and its lognormality principle. Other associations were observed between other PST metrics and visual/motor skills. These results suggest the potential of PST to identify children with suboptimal neuromotricity.

Keywords*—Prematurity, Neurodevelopmental Outcome, Kinematic Theory, Lognormality Principle, Pen Stroke Test on Digitizer.*

I. Introduction

One in ten live births occur prematurely <37 weeks of gestational age. The majority survives. Preterm birth takes place during a period of rapid brain growth and maturation. Neonates born prematurely can be exposed to perturbations related to preterm birth complications and treatments, such as a lack of oxygen, infections, malnutrition, cerebral bleeds, or steroids [1-4]. These conditions can alter normal brain maturational processes, which may later manifest as neurodevelopmental problems such as coordination and writing difficulties or attention-deficit/hyperactivity disorder (ADHD) [5, 6]. Children born preterm are more likely to face academic challenges due to problems with cognitive, language, motor, and behavioral skills that emerge during the preschool years [7-9]. 34-40% of children born preterm are not ready for school by the time they reach kindergarten [10]. Providing targeted early intervention services and addressing educational needs prior to school entry are essential to improve child functioning and quality of life [11-14]. Unfortunately, >50% of children with neurodevelopmental issues are not identified by primary care physicians [15-20].

To support primary care providers in identifying children with potential neurodevelopmental problems, screening tools can be useful. Such a tool yields a result indicating whether or not a problem that requires more comprehensive and specialized evaluation might be present. To be applicable in clinics, the tool needs to be easy and safe to administer, with adequate psychometric properties.

The *Pen Stroke Test on digitizer* (PST) from the *Lognometer toolbox* [21] is a tool based on handwriting production that may be used to assess typical and atypical motor development in children. It consists in producing handwriting strokes on a computerized interface which are then analyzed using the sigma-lognormal model [22, 23]. The sigma-lognormal model is considered to reflect a "perfect" motor control [24] and was validated in neuromonitoring studies [25, 26]. Using this model, the analysis of the velocity profile provides global parameters describing the general state of the neuromotor system and the quality of the modeling. Evidence supports the utility of the PST to differentiate levels of graphomotor performance in normal children at age 3-5 years [22]. Children with ADHD also differed from controls on PST metrics, suggesting its potential relevance as a screening instrument in various neurodevelopmental contexts [23]. Finally, the PST was used to detect a difference between handwriting strokes produced by children with mild traumatic brain injuries at 1 and 3 months post-injury [27].

The overall goal of this proof-of-concept study was to assess the PST as a developmental screening tool in preschool-aged children born prematurely. The objectives were to (1) evaluate the feasibility of using the PST in children born prematurely at age 4-5 years and (2) assess the association between PST metrics and motor proficiency and attention skills/behavioral regulation. We hypothesized that the PST can identify preschool-aged children born prematurely with suboptimal neurodevelopment.

II. Material & Methods

A. Study design and population

This is a single-center cross-sectional observational study. Children born prematurely between 29 and 36 weeks of gestational age were recruited at birth in the neonatal intensive care unit of the Centre Hospitalier Universitaire Sainte-Justine, Montréal, Canada. At age 4.5 years, participants underwent a neurodevelopmental evaluation by trained occupational therapists who were blinded to the screening test results. At the same time, the PST was conducted by another trained research assistant.

B. Study procedures

Motor proficiency was assessed by the Movement Assessment Battery for Children, 2nd Ed. (MABC-2). This standardized test assesses manual dexterity, aiming and catching, and balance (mean=10, standard deviation (SD)=3) [28]. A total score can be derived and determines the infant motor skills. The Beery-Buktenica Developmental Test of Visual-Motor Integration, 6th Ed. (VMI-6) was also used to assess visual perception, motor coordination and the integration of visual and motor abilities (mean=100; SD=15) [29]. Attention and behavioral regulation were evaluated with the Behavioral Assessment System for Children, 3rd Ed. - Parent Form (BASC3) [30].

The PST consisted of rapidly executing a set of three simple handwriting movements on a Lognometer, a digitizer (Wacom Cintiq 13HD, 200 Hz) connected to a laptop [21-23]. The first 2 movements consisted of making 15 rapid pen strokes, each time, between a starting zone, identified by a black point, and an arrival zone displayed in gray (Fig. 1A). For the first test, two green vertical bars are displayed for 500 ms on each side of the template as a visual GO signal while, for the second test, a sound cue (at 1kHz for 500 ms) is emitted by the computer. The third movement consisted in drawing a triangle 15 times by connecting three points displayed on the screen in response to the sound stimulus (Fig. 1B). For each movement, the experimenter showed the child how to do the task. The child was then asked to perform practice strokes. After data acquisition, signals were fitted to the sigma-lognormal model and optimal parameters were extracted using the in-house Lognometer software. These parameters provided central and peripheral information about the quality of the neuromotor control of the upper limb.

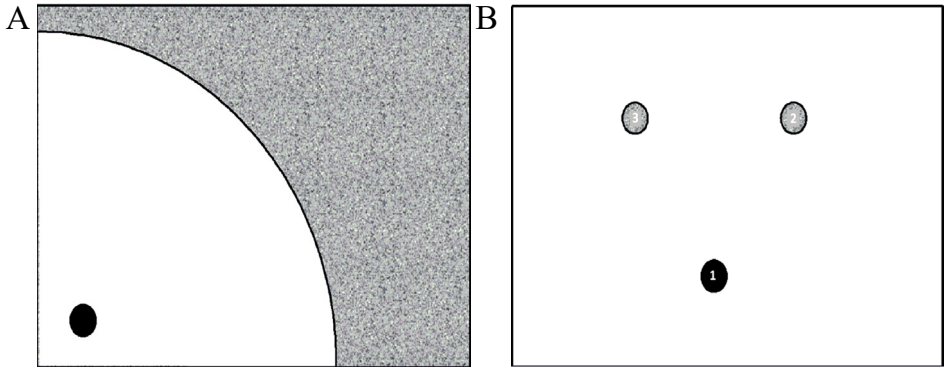


Fig. 1. (A) Starting zone (black dot) and arrival gray zone for the simple Pen Stroke Tests and (B) triangle template for the sequence of connecting strokes (1-2-3-1). For the visual tests, two vertical green bars (3.6 cm width) are displayed for 500 ms on each side of the 22 cm template.

The reconstruction quality was measured using 2 metrics: the signal-to-noise ratio (*SNR*) between the original and reconstructed velocity profile and the number of lognormal impulses (*NbLog*) required to reconstruct the stroke. *SNR* measured the reconstruction quality while *NbLog* the fluidity of the movement. In an ideal adult case, 2 lognormal impulses are required for the single stroke tests, while a minimum of four lognormal are required to produce triangles. The closer this number is to 2 or 4, the better the neuromotor ability. The time required for the brain to perform the motor command (t_0, a central parameter) and the time required to reach the median movement distance on a logarithmic scale (μ, a peripheral parameter) were also derived [25].

C. Statistical analysis

Descriptive statistics (counts and proportions, medians and interquartile ranges, means and standard deviations) were computed. Linear regression analyses were performed to assess the associations between PST metrics and neurodevelopmental outcome. Outputs included the correlation coefficient (r), 95% confidence interval (IC) and p-value. A significant association was defined as a $p < 0.05$.

III. Results

A. Population characteristics

In 2023-2024, 75 children underwent a neurodevelopmental assessment at median age 4.6 years (range 4.1-5.0). These children were born at a mean gestational age of 32.8 ± 1.8 weeks and 55% were males. Among infants, 29 (39%) displayed difficulties on manual dexterity, 19 (25%) on aiming and catching, and 42 (56%) on balance. 42 (56%) had suboptimal motor skills. On the Beery-Buktenica scales, 5 (6%) had suboptimal visuomotor abilities, 14 (19%) visual perception and 21 (28%) motor coordination. Overall, 13 (17%) had behavioral challenges on the BASC3.

B. Feasibility of the Pen Stroke Test

Among infants, 85% had usable data for the rapid pen strokes following a sound cue and 91% with the visual stimulus. However, we were not able to obtain valid data for the triangle task as majority of children were not able to trace the form.

C. Associations between the Pen Stroke Test and neurodevelopmental outcome

Fig. 2 and Fig. 3 shows associations between the PST metrics and neurodevelopmental outcome when an auditory cue was used. Fig. 2 shows linear associations between *SNR*, *NbLog* and t_0 with the MABC-2 manual dexterity and MABC-2 total scores: *SNR* was

Funding: The Canadian Institutes of Health Research and *La Fondation des Étoiles* (to M-NS, MD, and TML) as well The TransMedTech Institute (to RP).

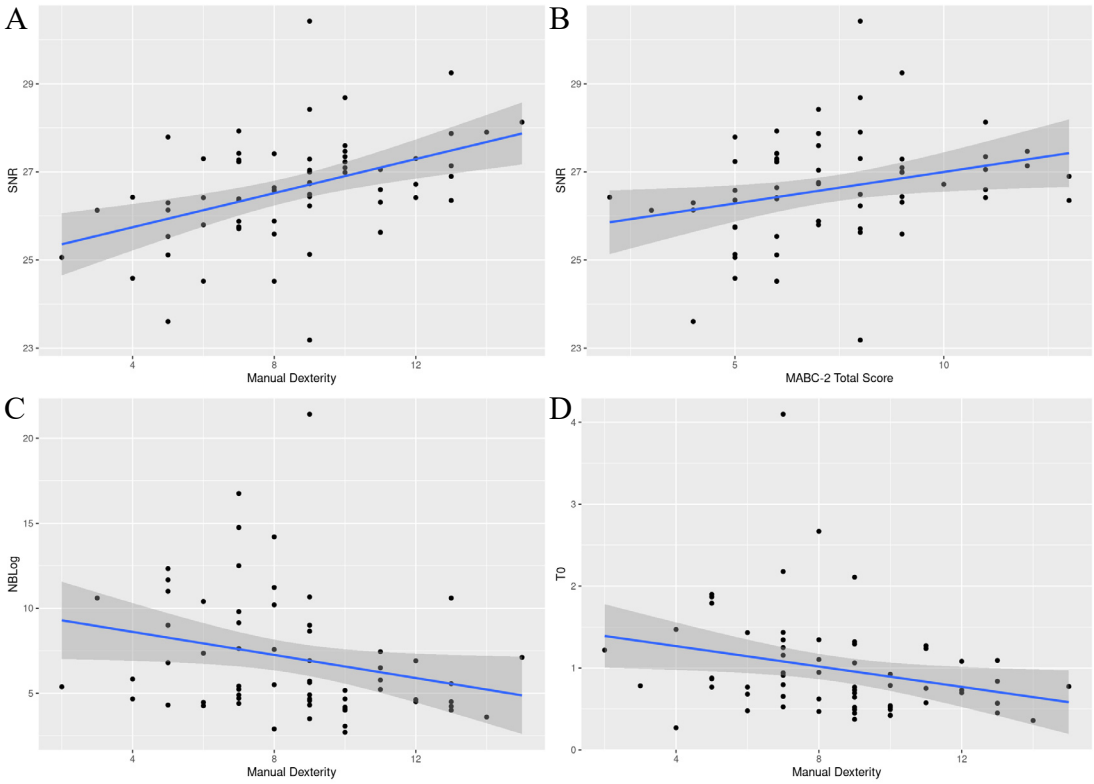


Fig. 2. Associations between the PST metrics (*SNR*, *NbLog* and t_0) with MABC-2 manual dexterity and MABC-2 total score.

positively associated with manual dexterity (Fig. 2A; $r = 0.44$, 95% CI = [0.22, 0.62], $p < 0.001$) and the total score (Fig. 2B; $r = 0.28$, 95% CI = [0.04, 0.49], $p = 0.025$) while *NbLog* (Fig. 2C; $r = -0.26$, 95% CI = [-0.47, -0.01], $p = 0.04$,) and t_0 (Fig. 2D; $r = -0.28$, 95% CI = [-0.49, -0.03], $p = 0.027$,) were negatively associated with manual dexterity. Fig. 3 shows linear associations between *NbLog* and t_0 with the Beery-Buktenica visual perception and motor coordination scores: *NbLog* and visual scores were negatively associated (Fig. 3A; $r = -0.26$, 95% CI = [-0.47, -0.01], $p = 0.04$,) while μ and motor scores were positively associated (Fig. 3B; $r = 0.25$, 95% CI = [0.01, 0.47], $p = 0.046$). All the other associations were non-significant including those between the PST and the BASC3. None of the associations were significant when a visual cue was used.

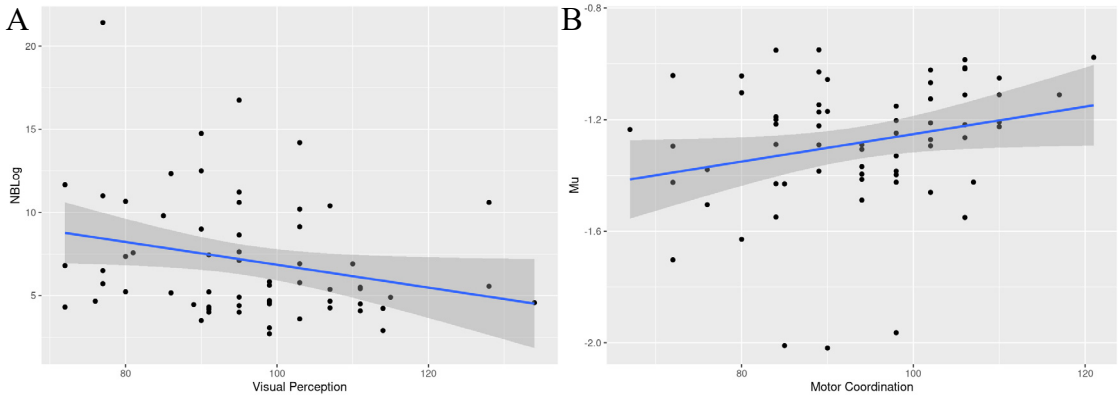


Fig. 3. Associations between the PST metrics (*NbLog* and μ) with Beery-Buktenica visual perception and motor coordination scores.

IV. Discussion & Conclusion

In this pilot study, metrics (SNR, NbLog, *t0*, and μ) derived from the PST were associated with neurodevelopmental (motor abilities and visual perception) outcome in preschool-aged children born prematurely. These associations suggest the potential role of the PST as a screening tool to identify suboptimal neurodevelopment in children born prematurely.

Higher *SNR* correlated with higher scores on manual dexterity and motor skills. This result is expected as *SNR* is a marker of the quality of the child pen stroke. Lower *NbLog* and t_0 also correlated with higher manual dexterity scores. This is consistent with the

kinematic theory and its lognormality principle [31] indicating that higher neuromotor ability on pen stroke corresponds to values of *NbLog* approaching a value of 2 (i.e., less impulses required to reach to reconstruct the stroke) [22]. Lower time required for the brain to perceive the stimulus and initiate the motor command (neuronal reaction time, t_0) was also associated with higher manual dexterity [25]. Lower *NbLog* was associated with better visual perception suggesting that children with good visual capabilities also displayed good neuromotor ability. Higher time required to reach the median movement distance on a logarithmic scale (μ) was associated with higher motor score, which may reflect a better motor integration. These results show the ability of the PST metrics to evaluate neuromotor development in children born prematurely and suggest its potential to identify children with suboptimal motor skills. Indeed, developmental delays can be the first manifestations of a permanent neurodevelopmental disorder, but it can also reflect immature brain development, especially in preterm populations. In both cases, it is important to identify these children since interventions can be initiated to allow catching up or to improve functional capacity.

Findings of this study are aligned with a previous assessment of PST metrics in children with and without ADHD [23]. In children with ADHD, motor planning (t_0) and fluidity (*NbLog*) were poorer compared to controls. Other correlations between the PST metrics with handwriting, manual dexterity, and information-processing and graphomotor speed were identified. In our group, only 17% of children had behavioral difficulties, which may explain the lack of associations between the PST metrics and behavioral scores. These findings support the use of graphomotor assessment in the clinical evaluation of children at risk of developmental difficulties.

In kindergarten children, our previous study showed that graphomotor criteria similar to the PST metrics allowed to differentiate children of different school grades [22]. This study demonstrated the sensitivity of these metrics in neurotypical children, which may help to develop strategy and preparation for handwriting learning at school.

There are several limitations in this pilot study. The sample size is modest (75 children) considering the wide range of gestational age at birth (29-36 weeks) and associated physiological differences. Stronger statistical effects on handwriting metrics may require a larger sample size. However, relevant and significant trends were identified in this study, which show the potential of the tool to identify children with suboptimal neurodevelopment. When using the visual cue as signal to perform the stroke, none of the associations between the Pen Stroke metrics with motor and behavioral assessments were significant. This inconsistency may be due to the type and duration of the stimulus (green vertical bars) used by the system, which may be insufficient to attract the attention of the child. Future technical improvements will be performed to assess the reaction of children when using other types of visual stimuli and durations. One potential extension of this work would be to compare our results with other kinematic variables that provide information on movement velocity and fluidity proposed in a previous study [32]. This comparison is beyond the scope of the present preliminary study. Inclusion criteria included a 48-hour admission in the neonatal intensive care unit, which could have resulted in selecting sicker neonates born between 29-36 weeks of gestational age. Thus, these findings are only applicable to this specific population, which is at higher risk of developmental delay.

Future developments include improvements in data processing procedures to reduce the number of stroke repetitions required, integration of rewards to increase the child interest, access to real-time analysis for point-of-care use, technical developments for telemedicine, and comprehensive psychometric assessment of the PST as a developmental screening tool. Indeed, the PST could eventually be used for large-scale neurodevelopmental screening of vulnerable children, thus reducing inequities in healthcare access.

In summary, the PST has the potential to assist health care professionals in identifying preschool-aged children born prematurely at greater risk of challenges with graphomotor skills, which is an important determinant of school readiness. The use of the PST would allow targeting children who would benefit from additional support and intervention to optimize learning.

References

[1] Boyle EM, Johnson S, Manktelow B, et al. Neonatal outcomes and delivery of care for infants born late preterm or moderately preterm: a prospective population-based study. *Arch Dis Child Fetal Neonatal Ed* 2015; 100: F479-485.

[2] Raju TN. Developmental physiology of late and moderate prematurity. *Semin Fetal Neonatal Med* 2012; 17: 126-131.

[3] Ramenghi LA. Late preterm babies and the risk of neurological damage. *Acta Biomed* 2015; 86 Suppl 1: 36-40.

[4] Vohr B. Long-term outcomes of moderately preterm, late preterm, and early term infants. *Clin Perinatol* 2013; 40: 739-751.

[5] Arpi E and Ferrari F. Preterm birth and behaviour problems in infants and preschool-age children: a review of the recent literature. *Dev Med Child Neurol* 2013; 55: 788-796.

[6] van Hoorn JF, Schoemaker MM, Stuive I, et al. Risk factors in early life for developmental coordination disorder: a scoping review. *Dev Med Child Neurol* 2021; 63: 511-519. 20201220.

[7] Allotey J, Zamora J, Cheong-See F, et al. Cognitive, motor, behavioural and academic performances of children born preterm: a meta-analysis and systematic review involving 64 061 children. *Bjog* 2018; 125: 16-25.

[8] Alterman N, Johnson S, Carson C, et al. Gestational age at birth and academic attainment in primary and secondary school in England: Evidence from a national cohort study. *PLoS One* 2022; 17: e0271952.

[9] Williams BL, Dunlop AL, Kramer M, et al. Perinatal origins of first-grade academic failure: role of prematurity and maternal factors. *Pediatrics* 2013; 131: 693-700.

[10] Louis D, Oberoi S, Ricci MF, et al. School Readiness Among Children Born Preterm in Manitoba, Canada. *JAMA Pediatr* 2022; 176: 1010-1019.

[11] Cioni G, Inguaggiato E and Sgandurra G. Early intervention in neurodevelopmental disorders: underlying neural mechanisms. *Dev Med Child Neurol* 2016; 58 Suppl 4: 61-66.

[12] Hughes AJ, Redsell SA and Glazebrook C. Motor Development Interventions for Preterm Infants: A Systematic Review and Meta-analysis. *Pediatrics* 2016; 138: e20160147.

[13] Litt JS, Glymour MM, Hauser-Cram P, et al. Early Intervention Services Improve School-age Functional Outcome among NICU Graduates. *Acad Pediatr* 2017; 18(468-474.

[14] Wallace IF, Berkman ND, Watson LR, et al. Screening for Speech and Language Delay in Children 5 Years Old and Younger: A Systematic Review. *Pediatrics* 2015; 136: e448-462.
[15] Guevara JP, Gerdes M, Localio R, et al. Effectiveness of developmental screening in an urban setting. *Pediatrics* 2013; 131: 30-37.
[16] Marks K, Hix-Small H, Clark K, et al. Lowering developmental screening thresholds and raising quality improvement for preterm children. *Pediatrics* 2009; 123: 1516-1523.
[17] Roberts G, Howard K, Spittle AJ, et al. Rates of early intervention services in very preterm children with developmental disabilities at age 2 years. *J Paediatr Child Health* 2008; 44: 276-280.
[18] Sheldrick RC, Merchant S and Perrin EC. Identification of developmental-behavioral problems in primary care: a systematic review. *Pediatrics* 2011; 128: 356-363.
[19] Thomas SA, Cotton W, Pan X, et al. Comparison of systematic developmental surveillance with standardized developmental screening in primary care. *Clin Pediatr (Phila)* 2012; 51: 154-159.
[20] Thomas R, Spragins W, Mazloum G, et al. Rates of detection of developmental problems at the 18 - month well - baby visit by family physicians' using four evidence - based screening tools compared to usual care: a randomized controlled trial. *Child Care Health Dev* 2016; 42: 382-393.
[21] O'Reilly C, Plamondon R and Faci N. The Lognometer : A New Normalized and Computerized Device for Assessing the Neurodevelopment of Fine Motor Control in Children. In: *26th International Conference on Pattern Recognition (ICPR)* 2022, pp.952-958.
[22] Duval T, Remi C, Plamondon R, et al. Combining sigma-lognormal modeling and classical features for analyzing graphomotor performances in kindergarten children. *Hum Mov Sci* 2015; 43: 183-200.
[23] Laniel P, Faci N, Plamondon R, et al. Kinematic analysis of fast pen strokes in children with ADHD. *Appl Neuropsychol Child* 2020; 9: 125-140.
[24] Plamondon R, O'Reilly C, Remi C, et al. The lognormal handwriter: learning, performing, and declining. *Front Psychol* 2013; 4: 945. 20131219.
[25] O'Reilly C, Plamondon R, Landou MK, et al. Using kinematic analysis of movement to predict the time occurrence of an evoked potential associated with a motor command. *Eur J Neurosci* 2013; 37: 173-180.
[26] Plamondon R, Djioua M and Mathieu PA. Time-dependence between upper arm muscles activity during rapid movements: observation of the proportional effects predicted by the kinematic theory. *Hum Mov Sci* 2013; 32: 1026-1039.
[27] Faci N, Désiré N, Beauchamp MH, et al. Lognormality in Children with Mild Traumatic Brain Injury : A Preliminary Pilot Study. In: Plamondon R, Marcelli, A., Ferrer, M.A. (ed) *Lognormality Principle and its Applications in e-Security, e-Learning and e-Health*. World Scientific Publishing Series in Machine Perception and Artificial Intelligence, 2021, pp.143-160.
[28] Brown T and Lalor A. The Movement Assessment Battery for Children--Second Edition (MABC-2): a review and critique. *Phys Occup Ther Pediatr* 2009; 29: 86-89.
[29] Beery KE, Beery NA and Buktenica NA. *Beery-Buktenica Developmental Test of Visual-Motor Integration - 6th Ed*. Bloomington, MN: PsychoCorp, 2010.
[30] Reynolds CR and Kamphaus RW. *Behavior assessment for children: Third edition.* Bloomington, MN: NCS Pearson, Inc., 2015.
[31] Plamondon R. The Lognormality Principle: a Personalized Survey. In: in Plamondon R, Marcelli, A.,Ferrer, M.A. (ed) *The Lognormality Principle and its Applications in e-Security, e-Learning and e-Health*. World Scientific Publishing Series in Machine Perception and Artificial Intelligence, 2021, pp.1-39.
[32] Danna J, Paz-Villagrán V and Velay JL. Signal-to-noise velocity peaks difference: a new method for evaluating the handwriting movement fluency in children with dysgraphia. *Res Dev Disabil* 2013; 34: 4375-84.

KNM2: A Collaborative Research Platform for Lifelong Lognormality Monitoring

Réjean Plamondon, Moussa Djioua, Lisa Jithame, Lina Sehad and, Rostand Marlone Njomo Njampou

Laboratoire Scribens, Génie Électrique, Polytechnique Montréal, 2500 Chemin de Polytechnique, Montréal, H3T-1J4, Québec, Canada
rejean.plamondon@polymtl.ca, m.djioua@polymtl.ca, lisa.jithame@polymtl.ca, lina.sehad@polymtl.ca,
rostand-marlone.njomo-njampou@polymtl.ca

***Abstract*—** Starting with a brief survey of the Kinematic Theory of human movements and its Lognormality Principle, this paper focuses on one of its fundamental predictions: the lognormality bridge. To test this hypothesis, the Kineneuromyometer (KNM2) platform is presented in the context of collecting data to build lognormality curves. Typical pen stroke samples are used to introduce the neuromotor growing segment of the bridge. A potential side effect of this paper is to foster international collaboration by exploring the possibility of launching a worldwide neuromotor data collection initiative.

Keywords—Kinematic Theory, Lognormality, KineNeuroMyoMetry, Sigma-Lognormal model, KNM2

I. INTRODUCTION

The Kinematic Theory of human movements celebrates its 30th anniversary this year [1,2,3,4] and as a perpetuity challenge, this paper presents a proposal for extending its impact among the community of researchers working in motor control. The whole paradigm relies on a few fundamental hypotheses:

1. The velocity is the control variable used by the Central Nervous System to produce reaching movement, because the velocity vector is always tangent to the trajectory, which allows learning from observations.

2. The invariant properties of single strokes reported for more than a century by numerous researchers [5,6,7] reflect the asymptotic behavior of complex systems, made up of a large number of coupled neuromuscular networks.

3. The Central Limit Theorem can be used to point out emergent phenomena in these complex systems. It predicts that the impulse response of a neuromuscular system is described by a lognormal velocity profile.

The physiological hypotheses of the Kinematic Theory have been validated through EMG [8] and EEG [9] experiments. It has been used and intensively tested by many researchers over the years [10,11], which has led to postulating the Lognormality Principle [10]: "The lognormal velocity patterns, which are the results of an asymptotic central limit convergence, can be interpreted as reflecting the behavior of subjects who are in perfect control of their movements" and as a corollary: "If we specifically focus on the basic mathematical convergence toward lognormality, movement control learning, on the one hand, can be interpreted as a migration toward an ideal representation of perfectly mastered movements. On the other hand, aging and health problems should reveal a progressive departure from lognormality. In between, human beings are taking advantage of lognormality." In other words, the production of complex movements is reached through the exploitation of this principle by time superimposing and summing up lognormal velocity vectors, with the goal of minimizing their number in each task, to produce efficient, fluent gestures optimizing the energy required to generate these.

II. LOGNORMALITY BRIDGE

One strong implication of the Lognormality Principle is that it predicts the existence of a universal lognormality bridge, see Fig.1. Indeed, from a motor control perspective, life can be seen as a bridge crossing. At any time, any living person can be positioned on this curve in terms of percentile as a function of his age. Once this person is positioned, the curve allows for the follow-up of his lognormality over the years. In this paper, we present the KineNeuroMyoMeter, KNM2 for short, a platform for collecting gesture data to build personalized lognormality bridges. In the next section 3, we describe the hardware, software, and network components of the system. In section 4, we introduce a typical data acquisition protocol in use at Laboratoire Scribens to collect and analyze such data. Typical handwritten gesture specimens are presented, as well as displaying a neuromotor growing curve for primary school kids aged between 6 and 12 years. These preliminary results are consistent with the growing segment of the bridge depicted in Fig.1, but more data is needed to cover the whole growth section. Similar studies are needed to build the performance and decline segment of the bridge. This cannot be done by a small group of researchers working on their own. In section 5, the conclusion sets the table for a worldwide neuromotor data collection initiative, with the goal of creating an international consortium of research partners willing to get involved in the challenge of building a universal lognormality bridge through data collection and sharing.

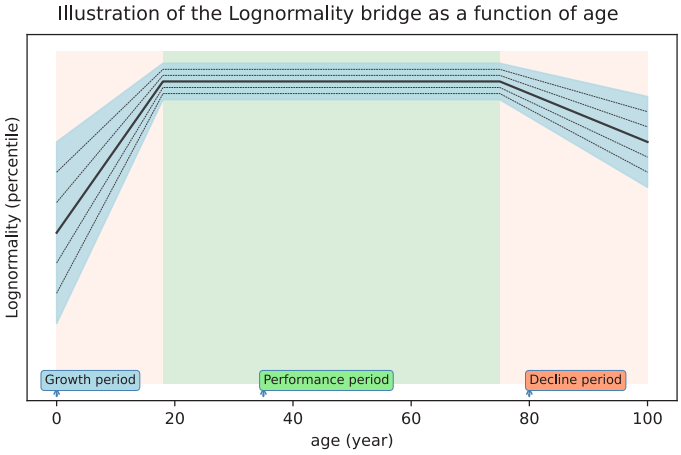


Fig.1: The lognormality bridge reflecting the three main phases of life: the growth, the performance, and the decline.

III. KNM2 A PLATFORM FOR COLLECTING DATA TO BUILT LOGNORMALITY BRIDGES

In its mission to characterize the lognormality of the global population, the Scribens Laboratory developed the KineNeuroMyoMeter (KNM2). This instrument measures fine motor dexterity of a person using Kinematic Theory of Human Movement and its Sigma-Lognormal model. The KNM2 is a mobile application installed on an iPad Pro, connected to the Scribens Laboratory server located at Polytechnique Montréal. Its function is to collect pen stroke data, send it through a secure connexion to the server for processing by proprietary software algorithms, and display, in real-time, measurements of person's fine motor anywhere in the world. Generally, data collection proceeds as follows: After registering through the app or website, the participant starts a data collection session with a non-identifiable anonymized code. Five psychophysical tests [13] are implemented and occur consecutively. The protocol involves performing 30 trials per task, with visual and auditory stimuli. Once the data collection is complete, the data are transmitted to and processed on the server, where they are securely stored in encrypted form in a separate database that contains no participant-identifying information; the results are then returned to the application for graphical display. The Fig.2 and Fig.3 illustrate both the global architecture of the KNM2, as well as the process of registering, recording stroke movements and the resulting display content.

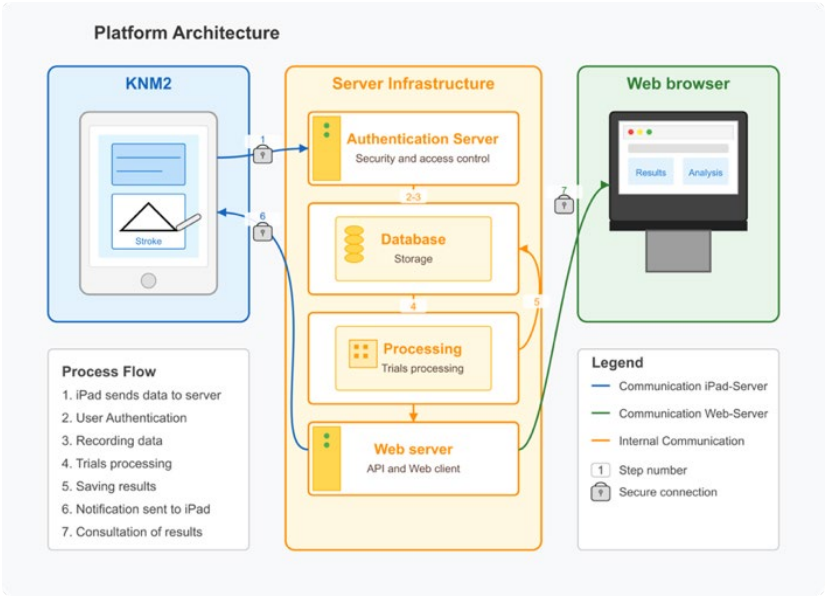


Fig.2: (a) the KNM2 global architecture.

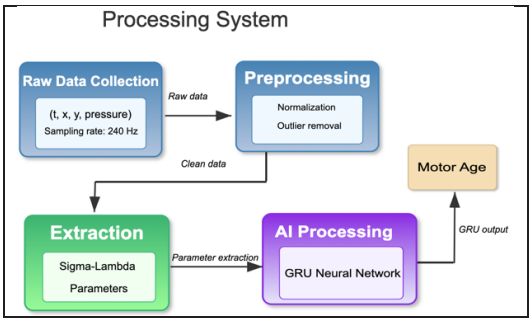


Fig.2: (b) the KNM2 processing system.

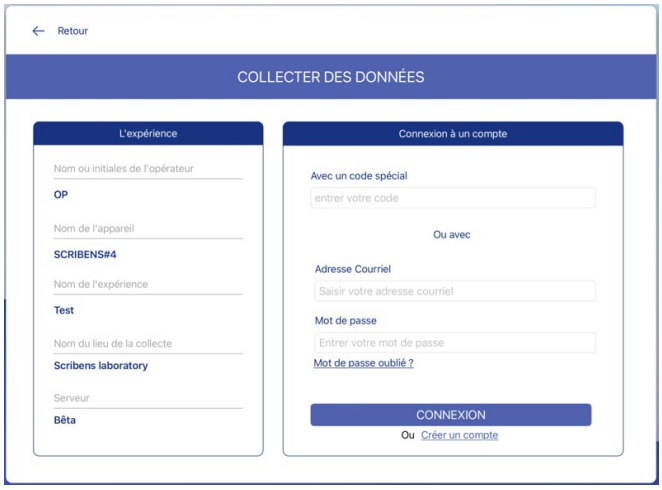


Fig.3: (a) KNM2 Login interface

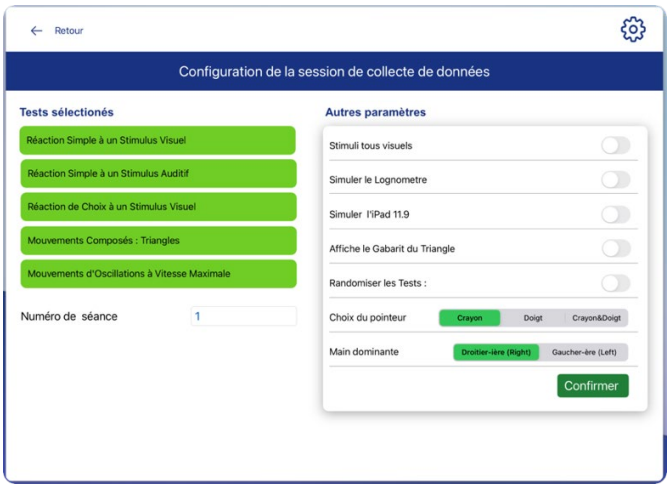


Fig.3: (b) KNM2 Tests Configuration Interface



Fig.3: (c) Typical pen stroke recorded using the KNM2

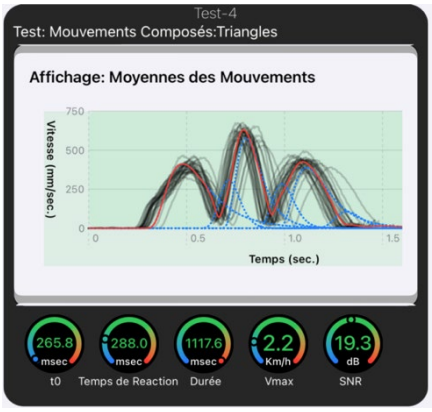
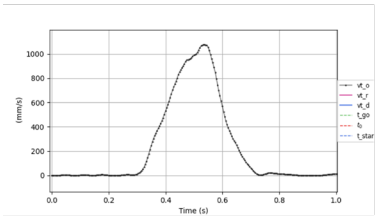


Fig.3: (d) Pen strokes visualization from KNM2

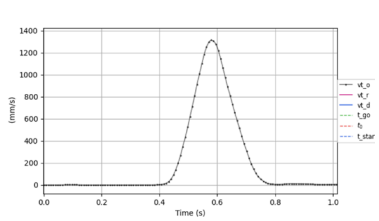
IV. Typical Specimens Samples and Example of the Neuromotor Growing Curve

The raw data and metadata are accessible for researchers to perform specific analyses and develop innovative signal processing and AI algorithms. Some samples of velocity profiles are depicted in Fig.4, as built from x,y trajectory coordinates.

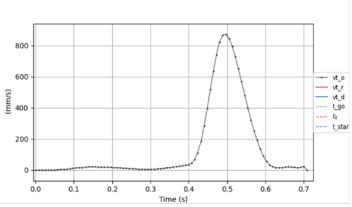
Furthermore, we developed a growth curve for children's motor development using the Lognometer, the predecessor of the KNM2 [12,13]. By analyzing movement kinematics, we estimated children's motor age with high precision. This standardized approach eliminates the subjectivity often found in traditional neurophysiological assessments, which rely on observer ratings. Unlike other motor skill tests that have been validated on small samples, the Lognometer was tested here on a large cohort of 500 children, allowing us to establish reliable benchmarks for fine motor development across 6 to 12 years old children [13,14], see Fig.5. This standardization enables us to classify children based on their percentile, much like growth curves that are used to track physical development. Ultimately, this tool will provide specialists with an objective method for assessing whether a child's motor skills follow a typical growth trajectory or point out potential developmental problems. Future research will focus on validating the tool's applicability across different cultural contexts to ensure its effectiveness in diverse populations. To further refine the growing segment of the bridge and fully leverage AI-driven analysis, additional data are needed, especially in age groups where the growing curve has gaps, from 3 to 6 and 12 to 18. Expanding our dataset will improve the precision of our models and enhance the tool's predictive capabilities. Additionally, the decline segment for older adults is under development, focusing on the natural decline in motor skills with aging. Overall, this would help to establish benchmarks for neuromotor aging and contribute to early detection strategies for age-related motor impairments. Regarding the performance segment, we have preliminary data from football athletes in their mid-twenties and young adults. This segment will have to be completed later as far as the lognormality is concerned, giving priority to growing and declining.



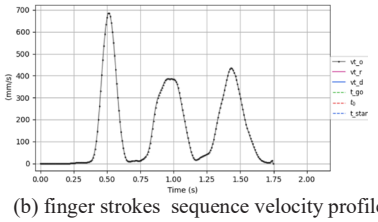
(a) pen stroke velocity profile of task 2 trial (visual simple reaction)



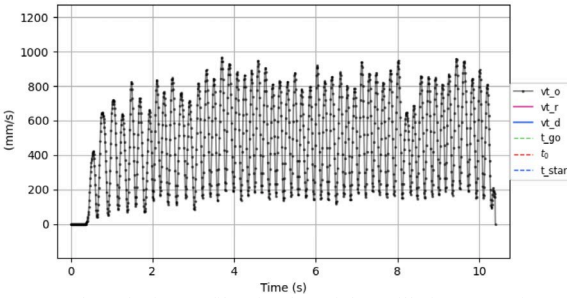
(b) finger stroke velocity profile of task 4 trial (auditory simple reaction)

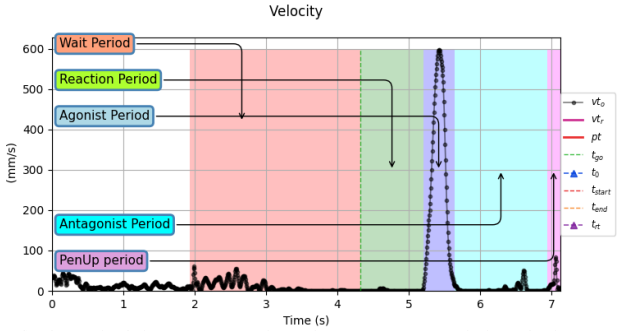


(a) pen stroke velocity profile of task 3 trial (visual choice reaction)



(b) finger strokes sequence velocity profile of task 6 trial (Triangle)





(e) pen stroke velocity profile of task 7 trial (oscillation at maximum speed) (f) Typical time segmentation of movement genesis in velocity space

Fig.4. Typical velocity profiles recorded from different motor tasks.

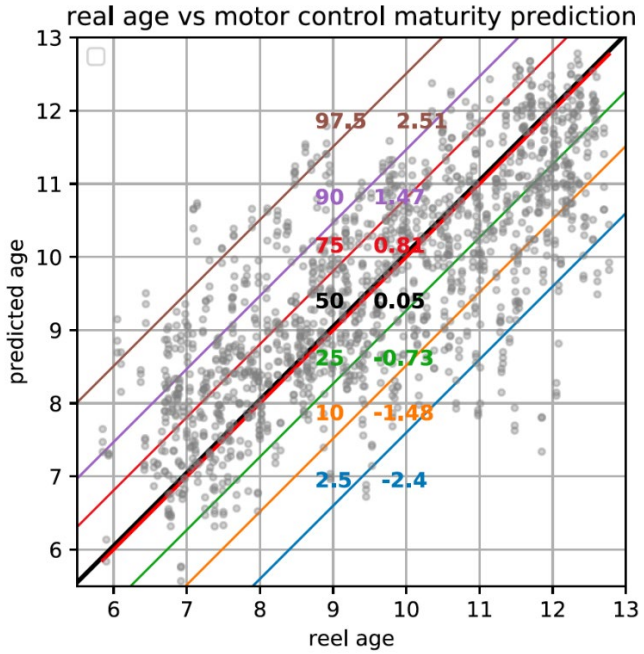


Fig 5. Neuromotor growing curve for children aged 6-12 [8]

V. Conclusion: a Worldwide Data Collection initiativee: looking for collaborators

The KNM2 is a platform to foster collaborative studies in human motor control through the Kinematic Theory and its Lognormality Principle. By enabling the collection and analysis of psychomotor data on a global scale, the KNM2 provides an opportunity to operationalize the universal lognormality bridge paradigm and characterize the neuromotor development, performance, and aging processes across different populations. Preliminary results from the present studies on children's motor development demonstrate the potential of this approach to create objective benchmarks for assessing fine motor skills and eventually pre-screening developmental problems. However, to fully realize the vision of a comprehensive neuromotor growth curve and extend this research to other age groups, particularly older adults, an international collaborative effort is essential. By expanding these datasets, various models could be refined to enhance predictive capabilities and contribute to early detection strategies for neuromotor impairments with various fine motor control biomarkers. We did such a move for handwritten characters and words in 1992 by creating the Unipen database [15] which boosted handwriting recognition performances. We should expect the same momentum if such a sharing project is initiated. Several questions will have to be addressed to reach this goal regarding the data acquisition protocol, the ethical issues in terms of data exchanges between countries, gate keepers as well as the financial, organizational and legal challenges to ensure the availability and sustainability of the data collected. As a follow up of IGS 2023 special session on lognormality [10], the IGS2025 conference offers a unique opportunity to contact researchers, clinicians, and institutions to join a worldwide data collection initiative and develop lognormality bridges. Through collective efforts, a robust database could be built to facilitate groundbreaking discoveries in human movement science, motor learning, and health diagnosis. By sharing data and expertise, standardized tools could be developed to benefit diverse populations and contribute to scientific advancements in neuromotor assessments. Together, under the auspices of the International Graphonomics Society (IGS – https://graphonomics.net/), we can create a powerful research network dedicated to understanding human movement with unprecedented precision, ultimately improving diagnosis, rehabilitation, and motor skill development worldwide. This is the optimistic take home message of this paper.

VI. Acknowledgments

This work was supported by Grant AXE-0113-POLY-AXELYS-GRANT from AXELYS, Ministère de l'Économie, de l'Innovation et de l' Énergie du Québec.

References

[1] Plamondon, R. "A kinematic theory of rapid human movements : Part I: Movement representation and generation". Biological Cybernetics..72:295-307, 1995

[2] Plamondon, R. "A kinematic theory of rapid human movements : Part II. Movement time and control". Biological Cybernetics. 72:309-320, 1995

[3] Plamondon, R. "A kinematic theory of rapid human movements : Part III. Kinetic outcomes". Biological Cybernetics..78:133-145, 1998

[4] Plamondon, R.,et al. "A kinematic theory of rapid human movement. Part IV: a formal mathematical proof and new insights". Biol. Cybern. 89, 126–138 2003.

[5] Bernshteĭn, N. A. *The Co-Ordination and Regulation of Movements*., Pergamon Press, 1967, http://books.google.com/books?id=F9dqAAAAMAAJ.

[6] Bizzi E, Mussa-Ivaldi FA, Giszter S. Computations underlying the execution of movement: a biological perspective. Science. 1991 Jul 19;253(5017):287-91. https://doi.org/10.1126/science.1857964. PMID: 1857964.

[7] Butchbach MER, Scott RC. Biological networks and complexity in early-onset motor neuron diseases. Front Neurol. 2022 Oct 21;13:1035406. https://doi.org/10.3389/fneur.2022.1035406. PMID: 36341099; PMCID: PMC9634177.

[8] Plamondon R, Djioua, M, Mathieu P.A. (2013). Time-Dependence Between Upper Arm Muscles Activity During Rapid Movements: Observation of the Proportional Effects Predicted by the Kinematic Theory. Human Movement Science. 32(5): 1026-1039.

[9] O'Reilly C, Plamondon R, Landou MK, Stemmers B. (2013). Using Kinematic Analysis of Movement to Predict the Time Occurrence of an Evoked Potential Associated with a Motor Command. European Journalof Neuroscience. 37(2): 173-180.

[10] Plamondon, R. (2021). The Lognormality Principle: a Personalized Survey.in Plamondon, R., Marcelli, A.,Ferrer, M.A.(Eds) The Lognormality Principle and its Applications in e-Security, e-Learning and e-Health. World Scientific Publishing Series in Machine Perception and Artificial Intelligence (88): 1-39.

[11] PLAMONDON R., BENSALAH A., LEBEL K., SALA-MEH R., SÉGUIN de BROIN G., O'REILLY C., BEGON M., DESBIENS O., GUY A., FOLLEMAYRIE F, BOYOGUÉNO-BIDIAS S-P., FISCHER A., ZHANG Z., ALAMARGOT D., RÉMI C., FACI N., FORTIN R., SIMARD M-N., BAZINET C. (2023). Lognormality: an open window on neuromotor control. Lecture Notes in Computer Sciences. Graphonomics in Human Body Movement: Bridging Research and Practice from Motor Control to Handwriting Analysis and Recognition. Proceedings 21st International Conference of the International Graphonomics Society, IGS 2023., Evora, Portugal (1205-258). Springer, Germany.

[12] Faci, N., Boyogueno-Bidias, S.P. Plamondon, R., Bergeron, N. (2021). An Interactive Tablet-based System to Run Neuromuscular Tests in Plamondon, R., Marcelli, A., & Ferrer, M. A. (Edst.) (2021). The Lognormality Principle and its Applications in e-Security, e-Learning and e-Health, World Scientific Publishing Series in Machine Perception and Artificial Intelligence (88): 269-288

[13] O'Reilly, C., Plamondon, R., Faci, N. (2022). The Lognometer : A New Normalized and Computerized Device for Assessing the Neurodevelopment of Fine Motor Control in Children. *2022 26th International Conference on Pattern Recognition (ICPR),* 952-058.

[14] Zhang, Z., O'Reilly, C., Plamondon, R. (2022). Comparing Symbolic and Connectionist Algorithms for Correlating the Age of Healthy Children with Sigma-Lognormal Neuromuscular Parameters. IIEEE proceedings of the 26th International Conference on Pattern Recognition (ICPR). pp. 4385-4391.

[15] Guyon I, Schomaker L, Plamondon R, Liberman M, Janet S. (1994). UNIPEN Project of On-Line Data Exchange and Recognizer Benchmarks. IEEE. Proc. 12th International Conference on Pattern Recognition, October 9-13, Jérusalem, Israël , pp 29-33.





Session

Posters

A Benchmark Dataset for Handwritten Document Layout Analysis in the Wild

Marina Gardella
Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, 91190 Gif-sur-Yvette, France
marina.gardella@ens-paris-saclay.fr

***Abstract*— Handwritten document layout analysis remains a challenging task due to the high variability in writing styles, page structures, and document degradations. Existing datasets often lack sufficient layout diversity, as they are typically sourced from homogeneous collections with similar structural patterns. This limitation hinders the development of robust models capable of generalizing to real-world scenarios. To address this issue, we introduce a new dataset of handwritten documents collected from Wikimedia Commons, representing a broad spectrum of historical and modern documents with varying layouts, languages, and writing conditions. Each document is annotated for layout analysis, with identified page segments and corresponding labels. While the dataset is not intended for large-scale model training, it serves as a valuable benchmark for evaluating layout analysis methods and identifying generalization challenges. By prioritizing layout diversity, this dataset provides a realistic testbed for advancing handwritten document segmentation and structural analysis, ultimately contributing to the development of more adaptable and reliable document processing systems. The dataset is publicly available for download at https://doi.org/10.5281/zenodo.15313600.**

Keywords*— *Handwritten Document Analysis, Document Layout Analysis, Benchmark Dataset, Page Segmentation

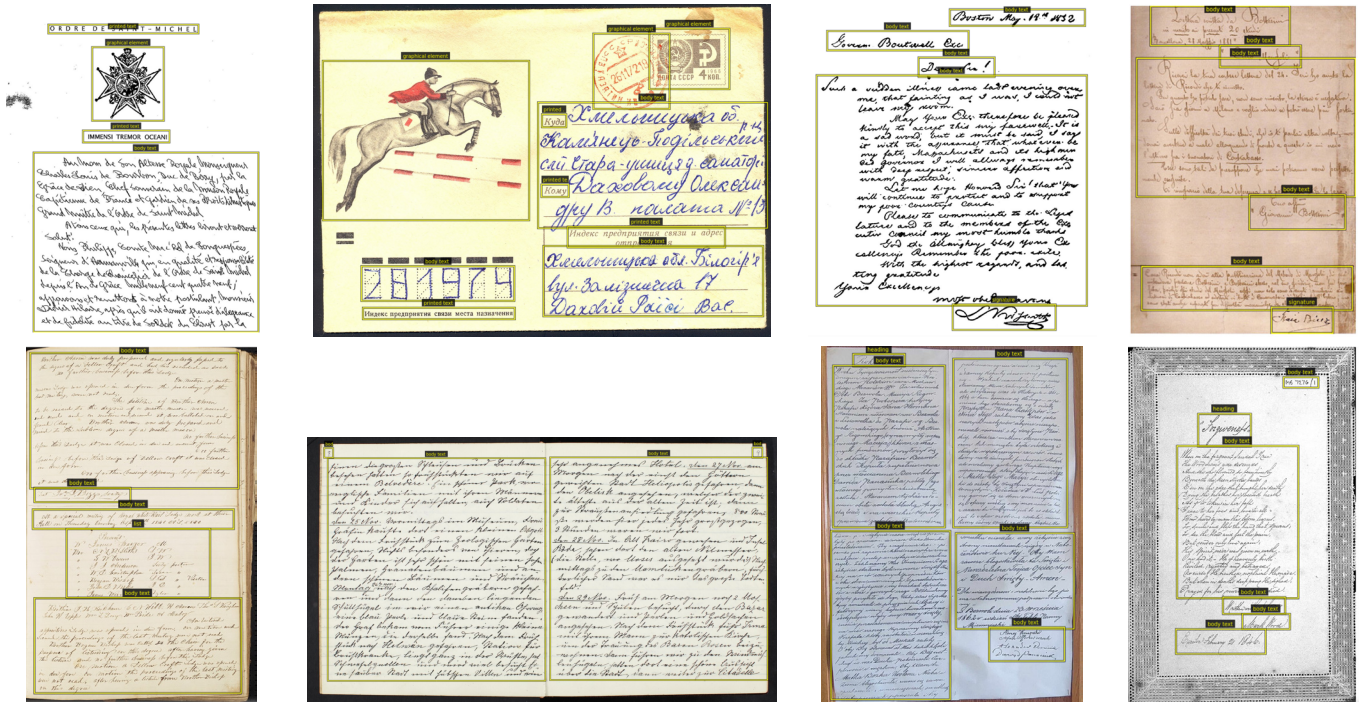


Fig. 1: Examples of document images in the proposed dataset showcasing the diversity in layout structures, handwriting styles, background textures, and document degradations. Each image includes annotations highlighting key structural elements such as body text, headings, graphical elements, signatures, printed text, lists and tables.

I. Introduction

Document layout analysis plays a crucial role in document image processing, enabling the extraction of textual and structural information from digitized pages [1]. It is essential for enabling downstream tasks like optical character recognition (OCR), content extraction, and layout reconstruction, ensuring that different elements are appropriately interpreted [2], [11]. While layout analysis for printed documents is well established, handwritten documents remain a major challenge due to their irregular structures, diverse writing styles, and the presence of marginal notes, overlapping text, and background noise. Developing robust algorithms for handwritten document layout analysis requires datasets that accurately reflect this real-world variability. However, most existing datasets for handwritten layout analysis suffer from a significant limitation: Their layouts tend to be

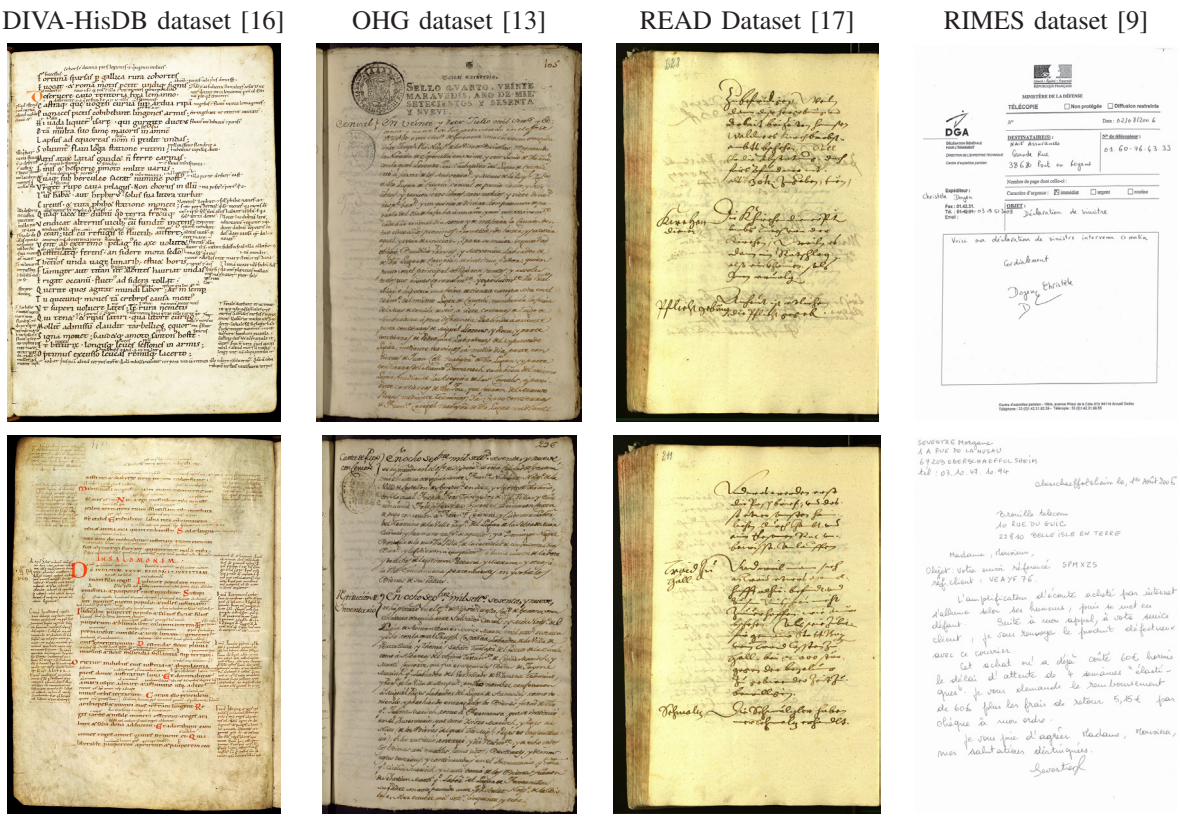


Fig. 2: Examples of document images in publicly available datasets DIVA-HisDB dataset [16], OHG dataset [13], READ Dataset [17] and RIMES dataset [9]. Note that, within each dataset, documents are pretty homogeneous in style, structure and condition.

highly homogeneous within the dataset. Many collections focus on specific document types, such as historical manuscripts from a single source [13], [16], [17] or handwritten forms with predefined structures [9]. As a result, models trained on these datasets often struggle to generalize to more diverse real-world scenarios where document layouts vary widely. This lack of variability hinders progress in building adaptable and resilient layout analysis systems.

To address this gap, we introduce a new dataset of handwritten documents collected from Wikimedia Commons. Using publicly available and heterogeneous sources, our dataset encompasses a diverse range of document layouts, spanning different historical periods, languages, writing styles, and levels of degradation, as illustrated in Figure 1. Each document is carefully annotated for layout analysis, with identified page segments and their corresponding labels. Although the data set is relatively small and is not intended for model training, it serves as a valuable benchmark for evaluating layout analysis methods and identifying generalization challenges. By prioritizing layout diversity, we aim to support the development of more robust handwritten document analysis techniques. We hope that this resource will aid in improving segmentation algorithms and enhancing generalization across various handwriting styles and document structures, ultimately advancing the field of handwritten document understanding.

II. Related works

Handwritten document layout analysis relies heavily on high-quality annotated datasets to develop and evaluate segmentation algorithms. While several widely used datasets have significantly contributed to t he field, many exhibit limited variability in layout structure, language, script, or source, which can hinder the evaluation of layout analysis systems intended for real-world, heterogeneous documents. Figure 2 illustrates some examples of these datasets, underscoring the relatively uniform visual and structural patterns they contain.

The READ Dataset [17], derived from the Ratsprotokolle collection of 15th–19th century German council meeting minutes, features challenging handwriting but limited diversity in document structure, degradation, and writing style. Its homogeneity makes it difficult t o a ssess h ow w ell m odels g eneralize b eyond t he n arrow c haracteristics o f t his d ataset. S imilarly, t he RIMES

dataset [9] captures handwriting variability by soliciting letters from volunteers, but the documents consistently follow a personal letter format, resulting in uniform layout structures. This restricts its usefulness for evaluating models on structurally diverse handwritten material. The OHG dataset [13], consisting of Spanish notarial deeds from the 18th century, is another valuable historical collection that suffers from layout homogeneity. The documents largely conform to legal formats with repetitive structure, which can bias models toward recognizing formulaic layout patterns rather than adapting to broader variability. The BH2M dataset [7] consists of 174 pages of Old Catalan marriage records written by a single scribe. While valuable for semantic and structural analysis in archival documents, the consistent authorship and single document type introduce potential bias and restrict structural variety. Similarly, HisClima [14] presents a dual-mode collection of 419 ship log pages (some with tables, others with narrative entries), but all documents are drawn from a narrow historical and functional domain.

The ICDAR19 HDRC Chinese dataset [15] contains over 11,000 historical Chinese family record pages. Although the dataset is large and linguistically rich, its focus on a single document type may introduce structural bias, and the strong cultural and script-specific nature of the content may limit transferability to other handwritten domains. The FCR dataset [12], comprising 500 pages of Finnish 19th-century court records, includes complex layouts with both single and double-page formats and annotated regions such as paragraphs, marginalia, and tables. It provides a rare example of heterogeneous, formal layout structures, but its domain specificity and language restriction may still limit generalization to more informal or cross-domain handwriting.

The DIVA-HisDB dataset [16] provides detailed layout annotations of medieval manuscripts, but it includes only three distinct books, leading to limited diversity in page structure and visual style. The HBA dataset [10], built from the Gallica digital library, includes 4,436 pages across 11 books spanning six centuries. It provides pixel-wise ground truth for six semantic layout classes, including handwritten and printed text. Although it offers significant variety in time period, script, and typography, its layout annotations are coarse and class-driven, which may be limiting for fine-grained structural analysis or hierarchical segmentation tasks. The IAM-HistDB [8] includes three popular sub-datasets: Saint Gall, Parzival, and George Washington (GW), containing historical Latin, Old German, and English texts respectively. While these datasets are widely used for handwriting recognition and word spotting, they are relatively small in number of pages and constrained in terms of layout complexity. Saint Gall and Parzival offer high-quality binarized and normalized text-line images, but their original layouts derive from single genres and are often authored by one or a few individuals. GW documents offer slightly more diversity in script and writer but still follow a formal correspondence structure. These datasets provide valuable ground truth for word and line-level tasks but are limited in capturing the diversity of modern or cross-cultural handwritten layouts.

The HORAE dataset [3] offers 557 pages selected from a large corpus of 500 manuscripts. The selected pages span multiple layout categories, including full-page text, miniatures, and marginalia. While the annotation process deliberately includes both typical and outlier layouts, the selection relies on automated clustering and sampling, and only two pages were chosen per layout class. Thus, while diverse in appearance, the dataset may not fully capture intra-class variability or offer enough samples per category for robust training and evaluation. Another of the few datasets that addresses layout complexity is the cBAD dataset [4], developed for the ICDAR 2017 competition. It includes skewed text lines, marginal notes, and challenging structures. However, its annotations focus on text-line detection, not full region-level layout segmentation, limiting its applicability for comprehensive layout analysis research.

III. The dataset

A. Document sources

The documents in our dataset were collected from Wikimedia Commons, an open-access repository that hosts a vast collection of historical and modern handwritten documents contributed by libraries, archives, and individual users. To construct our dataset, we selected a subset of handwritten documents that exhibit significant layout variability. This includes materials with different text arrangements, such as densely written pages, documents with marginalia, multi-column layouts, graphical elements, and irregularly spaced text blocks. The selection process prioritized diversity in page structure, aiming to create a dataset that better represents real-world challenges in handwritten document layout analysis. All collected documents are freely available under open licenses, ensuring that our dataset can be distributed and used for research without legal restrictions. The source of each individual document image is available together with the dataset.

B. Dataset composition

The dataset comprises 55 document images, representing a diverse selection of handwritten materials. While some images may have undergone JPEG-compression before being sourced, all files in the dataset are provided in PNG format to preserve quality. No preprocessing was applied to the document images, ensuring that they retain their original characteristics, including potential degradations, background noise, and artifacts. This decision was made to maintain the authenticity of the documents and better simulate real-world challenges in handwritten document layout analysis. The images exhibit considerable variation in resolution. On average, images measure 941×1130 pixels, with a standard deviation of 355×353 pixels. The smallest image in the dataset is `30.png` (580×372 pixels), while the largest is `06.png` (1939×3022 pixels), highlighting the diversity of

Label	Description
`body text`	The primary text content of the document, typically uniform in font size and style. It also includes any text not falling into other categories, such as page numbers.
`heading`	Text that appears in a larger or distinct font, usually marking the start of a section or the document itself.
`graphical element`	Includes figures, stamps, seals, and other non-textual elements. Decorative page ornaments are excluded, as they are considered part of the background.
`signature`	Personalized text, often in a distinct font from the body text, typically incorporating unique flourishes or strokes. Only signatures in a different font than the body text are labeled as such.
`printed text`	Any text printed (not handwritten), such as machine-printed characters or typeset content.
`list`	A series of items presented in a specific order, typically arranged vertically. Items may not necessarily span the full length of the page.
`table`	Structured text organized into rows and columns, creating a grid-like format.

TABLE I: Detailed criteria for assigning labels in the proposed dataset.

document formats. File sizes range from 55.49 KB (`37.png`) to 1454.56 KB (`19.png`), with an average of 327.94 KB. The total size of the dataset amounts to approximately 18.5 MB. Most images are in RGB color mode (45 files), while a smaller subset are in grayscale (8 files) or RGBA (2 files). This variety in color encoding and resolution reflects real-world scanning practices and introduces additional variability that layout analysis systems must handle. No resizing, cropping, or conversion was applied, reflecting the heterogeneous nature of real-world handwritten documents.

C. Annotations

Document images were manually annotated using the *VGG Image Annotator* (VIA) [5], [6]. VIA is a lightweight, browser-based annotation tool developed by the Visual Geometry Group, requiring no installation and supporting offline use, making it a convenient choice for manual document annotation. Region boundaries are defined using bounding boxes and, to each of them, a label is assigned. The possible labels are: `body text`, `heading`, `graphical element`, `signature`, `printed text`, `list` and `table`. Table I describes the criteria used to assign such labels.

D. Layout Structure Analysis

We analyzed the spatial and size characteristics of annotated regions across the handwritten document dataset. For each region, we computed its area as a percentage of the page area, and normalized the coordinates of its geometric center (X and Y) to the range $[0, 1]$ based on the page dimensions. The summary statistics in Table II report the number of instances per label, the mean relative area, and the average spatial position along with the corresponding position biases. Standard deviation values are also included to reflect the variability of region sizes across documents. The most frequent category, `body text`, accounts for 168 regions with a mean area of 15.79% of the page, though the standard deviation indicates considerable variability in their extent across pages. On average, these regions are centered both horizontally and vertically, indicating no spatial preference along either axis. Printed text occupies a minimal average space (0.54%) with relatively low variability. This is expected, as the document images were selected specifically for their dominance of handwritten text, with printed text being secondary. As expected, regions labeled as `heading` are positioned toward the top of the page (mean $\mathrm{Y} = 0.18$), with moderate area and standard deviation values. Visually prominent components such as graphical elements, lists, and tables display large average areas (ranging from 18.0% to 57.7%), with significant variability, suggesting that their size and placement can differ substantially depending on document structure. Regions corresponding to the `signature` label are compact (mean area 3.77%) and consistently located in the lower-right area of the page (mean $\mathrm{X} = 0.66$, mean $\mathrm{Y} = 0.92$), in line with typical document conventions.

IV. Concluding remarks

While significant efforts have been made to generate large-scale annotated datasets for handwritten document analysis, less attention has been devoted to evaluating the generalization abilities of the existing methods. In particular, many existing datasets exhibit limited variability in terms of layout diversity, which restricts their ability to train models that generalize well to real-world scenarios. The dataset introduced in this paper aims to address this gap by providing a collection of handwritten documents that capture a wide range of historical periods, languages, handwriting styles, and degradations. Although our dataset is smaller in scale compared to others, its reduced size offers a unique advantage—enabling visual inspection and qualitative analysis of the results. This approach allows for a more nuanced understanding of how layout analysis methods perform under varying conditions and reveals potential generalization issues that might not be evident with larger, less diverse datasets. We believe that this resource will contribute to the development of more robust layout analysis techniques, encouraging

Label	Count	Mean Area (%)	Std Area (%)	Mean X	Mean Y	Position Bias
body text	168	15.79	22.65	0.53	0.45	center-center
graphical element	14	16.79	25.07	0.55	0.35	top-center
heading	22	4.57	5.25	0.51	0.14	top-center
list	7	40.69	20.0	0.54	0.51	center-center
printed text	54	0.45	0.57	0.47	0.63	bottom-center
signature	10	2.99	2.13	0.63	0.81	bottom-right
table	5	43.63	26.75	0.55	0.42	center-center

TABLE II: Summary statistics for each region label: count, relative area (mean and standard deviation), normalized center positions, and spatial bias indicating positional tendencies.

further exploration into the evaluation of model generalization. Ultimately, we hope that this dataset will support the continued advancement of handwritten document understanding and the improvement of segmentation and analysis algorithms across a broader range of real-world applications.

References

[1] Henry S. Baird. Anatomy of a page reader. In *Proceedings of the 2nd IAPR Workshop on Machine Vision Applications (MVA) 1990, November 28-30, 1990, Kokubunji, Tokyo, Japan*, pages 483–486, 1990. https://dblp.org/rec/conf/mva/Baird90.html.

[2] Showmik Bhowmik, Soumyadeep Kundu, and Ram Sarkar. BINYAS: a complex document layout analysis system. *Multimedia Tools and Applications*, 80:1–34, 03 2021. https://doi.org/10.1007/s11042-020-09832-3.

[3] Mélodie Boillet, Marie-Laurence Bonhomme, Dominique Stutzmann, and Christopher Kermorvant. HORAE: An Annotated Dataset of Books of Hours. In *Proceedings of the 5th International Workshop on Historical Document Imaging and Processing (HIP)*, pages 7–12, Sydney, NSW, Australia, 2019. Association for Computing Machinery. https://doi.org/10.1145/3352631.3352633.

[4] Markus Diem, Florian Kleber, Stefan Fiel, Tobias Grüning, and Basilis Gatos. cBAD: ICDAR2017 competition on baseline detection. In *Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 01, pages 1355–1360, 2017. https://doi.org/10.1109/ICDAR.2017.222.

[5] Abhishek Dutta, Ankush Gupta and Andrew Zisserman. VGG Image Annotator (VIA). 2016. Version: 2.0.12, Accessed: 3 Mars 2025. https://www.robots.ox.ac.uk/~vgg/software/via/

[6] Abhishek Dutta and Andrew Zisserman. The VIA annotation software for images, audio and video. In *Proceedings of the 27th ACM International Conference on Multimedia (MM)*, New York, NY, USA, 2019. ACM. https://doi.org/10.1145/3343031.3350535.

[7] David Fernández-Mota, Jon Almazán, Núria Cirera, Alicia Fornés, and Josep Lladós. BH2M: The Barcelona Historical, Handwritten Marriages Database. In *Proceedings of the 22nd International Conference on Pattern Recognition (ICPR)*, pages 256–261, 2014. https://doi.org/10.1109/ICPR.2014.53.

[8] Andreas Fischer, Markus Wuthrich, Marcus Liwicki, Volkmar Frinken, Horst Bunke, Gabriel Viehhauser, and Michael Stolz. Automatic Transcription of Handwritten Medieval Documents. In *Proceedings of the 15th International Conference on Virtual Systems and Multimedia (VSMM)*, pages 137–142, 2009. https://doi.org/10.1109/VSMM.2009.26.

[9] Emmanuèle Grosicki, Matthieu Carre, Jean-Marie Brodin, and Edouard Geoffrois. RIMES evaluation campaign for handwritten mail processing. In *Proceedings of the 11th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 1 – 6, Montreal, Canada, August 2008. Concordia University.

[10] Maroua Mehri, Pierre Héroux, Rémy Mullot, Jean-Philippe Moreux, Bertrand B. Coüasnon, and Bill Barrett. HBA 1.0: A Pixel-based Annotated Dataset for Historical Book Analysis. In *International Workshop on Historical Document Imaging and Processing (HIP)*, Kyoto, Japan, 11 2017. https://doi.org/10.1145/3151509.3151528.

[11] George Nagy and Sharad C. Seth. Hierarchical representation of optically scanned documents. In *Proceedings of the 7th International Conference on Pattern Recognition (ICPR)*, pages 347–349, 1984.

[12] Lorenzo Quirós, Maria Kallio, and Enrique Vidal. Finnish Court Records-sub500. A dataset of Finnish notarial records (19th Century). 2020. https://doi.org/10.5281/zendo.3945087.

[13] Lorenzo Quirós, Lluis Serrano, Vicente Bosch, Alejandro H. Toselli, Rosa Congost, Enric Saguer, and Enrique Vidal. *Oficio de Hipotecas de Girona. A dataset of Spanish notarial deeds (18th Century) for Handwritten Text Recognition and Layout Analysis of historical documents.* 2018. [Data set]. https://doi.org/10.5281/zenodo.1322666.

[14] Verónica Romero and Joan Andreu Sánchez. The HisClima database: historical weather logs for automatic transcription and information extraction. In *Proceedings of 25th International Conference on Pattern Recognition (ICPR)*, pages 10141–10148, 2021. https://doi.org/10.1109/ICPR48806.2021.9412210.

[15] Rajkumar Saini, Derek Dobson, Jon Morrey, Marcus Liwicki, and Foteini Simistira Liwicki. ICDAR 2019 Historical Document Reading Challenge on Large Structured Chinese Family Records. In *Proceedings of the 15th International Conference on Document Analysis and Recognition (ICDAR)*, pages 1499–1504, 2019. https://doi.org/10.1109/ICDAR.2019.00241.

[16] Foteini Simistira, Mathias Seuret, Nicole Eichenberger, Angelika Garz, Marcus Liwicki, and Rolf Ingold. DIVA-HisDB: A Precisely Annotated Large Dataset of Challenging Medieval Manuscripts. In *Proceedings of the 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 471–476, 2016. https://doi.org/10.1109/ICFHR.2016.0093.

[17] Joan Andreu Sánchez, Verónica Romero, Alejandro H. Toselli, and Enrique Vidal. ICFHR2016 Competition on Handwritten Text Recognition on the READ Dataset. In *Proceedings of the 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 630–635, 2016. https://doi.org/10.1109/ICFHR.2016.0120.

A Survey on Skeletonization of Character Images

Yunqi Xu
CENPARMI
Gina Cody School of Engineering and Computer Science
Concordia University
Montreal,Quebec, Canada
x_yunq@encs.concordia.ca

Ching Y. Suen
CENPARMI
Gina Cody School of Engineering and Computer Science
Concordia University
Montreal,Quebec, Canada
suen@encs.concordia.ca

***Abstract*— In handwriting analysis, skeleton extraction is a critical preprocessing step that standardizes input data and preserves essential shape features. Early thinning algorithms, such as Zhang-Suen, were widely adopted for their simplicity. Recent research has demonstrated the potential of deep learning approaches—including CNN-based and GAN-based architectures—to enhance robustness, adapt to varying styles, and address noise and degradation issues. These advancements have improved accuracy in handwriting recognition, calligraphy font generation, and structural analysis, thereby advancing both practical applications and theoretical development. In this paper, we present a comprehensive survey of skeleton extraction algorithms developed in recent years.**

Keywords*— *Graphology analysis, Handwriting analysis, Skeleton extraction, Thinning algorithm, Deep learning

I. Introduction

Skeleton extraction is a key preprocessing step in handwriting analysis and likewise underpins graphology—an applied branch that probes human personality and stylistic traits. An ideal skeleton should be free of redundant pixels and extraneous branches, remain distortion-free, and faithfully capture the original character's topology. Although humans can intuitively discern a character's skeleton—even in unfamiliar scripts—machines often struggle to achieve comparable precision.

Thinning algorithms fall into three main categories: sequential, parallel, and medial axis methods. Sequential methods remove contour pixels iteratively in a fixed order, which can lead to anisotropic outputs. Parallel methods improve upon this by removing pixels simultaneously based on prior iterations, yet they remain susceptible to noise and heuristic dependence. In contrast, medial axis methods employ distance transforms to extract central lines, offering enhanced noise robustness though sometimes introducing minor distortions at corners and junctions.

Skeleton extraction (thinning) preserves essential structural information for shape recognition and stroke extraction while reducing data volume, enhancing the extraction of distinctive features, and accelerating digital processing. Consequently, integrating skeletonization into processing pipelines significantly improves the performance of character image reconstruction and recognition, with these benefits extending to applications such as fingerprint identification, biometric authentication, signature verification, and sketch-based image retrieval.

In the remainder of this paper, we first review several well-known skeleton algorithms in handwriting analysis (Section II). Next, we introduce new skeleton methods, ranging from traditional computer vision approaches III-A to the latest deep learning techniques III-B (Section III), and the application on skeleton algorithm III-C. Finally, we discuss potential future research directions in Section IV.

II. Review of traditional methods

In this section, we first review several traditional thinning algorithms and examine how recent studies leverage them for feature extraction in tasks such as restoration, recognition, and synthesis. Thinning—also referred to as skeletonization—has a long history in image processing, particularly in handwriting analysis. The seminal Zhang-Suen [1] algorithm introduced in 1984 iteratively reduces binary images to single-pixel-wide skeletons while preserving connectivity. Following, various well-known approaches have been proposed, including neighborhood-based strategies that iteratively remove boundary pixels until only central lines remain [2], [3], as well as distance-based methods relying on distance transforms to directly identify the medial axis [4], [5]. Meanwhile, [6] highlighting key differences between skeletonization algorithms (Zhang-Suen [1], Guo-Hall [7], and Lee-Kashyab-Chu [8]), and [9] highlights human coherence in comparing skeletons and the novelty of using reference skeletons to facilitate the evaluation of thinning algorithms. A broader overview can be found in [10]. Collectively, these works underscore the importance of skeleton extraction in normalizing complex degradations, enhancing feature representation, and enabling diverse document analysis and computer vision applications.

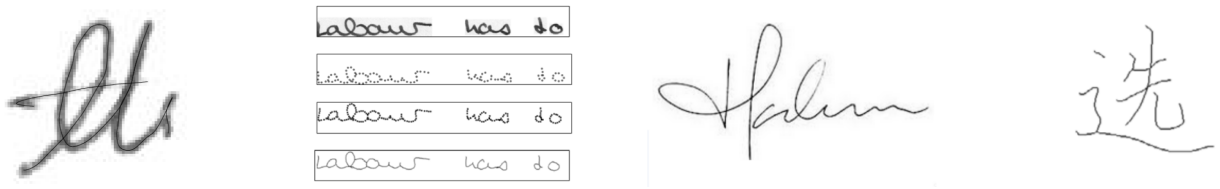
III. New Methods

Recent years, the skeleton algorithms are no more treated as a single research direction but usually connected with other method. The following part will introduce some new methods in the past decades.

A. Method based on traditional calculation

In [11], grayscale images are skeletonized by detecting stroke centerlines, representing them with cubic B-splines, and refining those splines via a genetic algorithm; a final fine-tuning step aligns near-junction points. Building on this idea of precise skeleton extraction, [12] unifies offline and online handwriting by converting both into consistent skeleton images, using a proprietary stroke extractor for offline data, baseline and slant normalization, and Bezier-based rendering to produce single-pixel-wide strokes. A different strategy is adopted in [13], where Growing Neural Gas (GNG) approximates the character's topology before a Relative Neighborhood Graph (RNG) removes extraneous edges. In contrast, [14] introduces a loop-based skeleton model to capture pen-down and pen-up events for more detailed recognition and forensic analysis.

Meanwhile, [15] applies a ring radius transform (RRT) to extract the medial axis from Canny edges, reconnects disjoint segments via an iterative-maximal-growing (IMG) algorithm, and removes false axes with color-based clustering. After a modified iterative-midpoint (IMM) method that fills remaining gaps, the skeleton is processed by Tesseract, achieving strong thinning and recognition performance. To further enhance skeletonization and trajectory recovery, [16] combines triangulation and graph theory for near-human handwriting output. In [17], a graph-based skeleton model and ambiguous zone analysis guide recovery of multi-stroke drawing order, while [18] (the SCAD method) extracts a single-pixel-wide skeleton, removes burrs, identifies ambiguous zones, and reconstructs missing segments by interpolation, attaining 98.37% accuracy for hard-pen regular script. Fig. 2 presents some result from our selected papers.



(a) [11] (b) [12] (c) [14] (d) [18]

Fig. 1: Some Results From Selected Papers

B. Deep Learning Methods

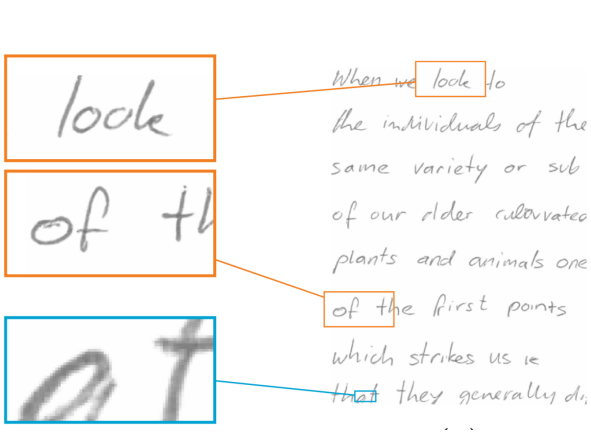
While the previous section focused on traditional skeleton extraction algorithms rooted in rule-based heuristics or geometry-based computations, the following section explores emerging deep learning approaches that leverage neural architectures for enhanced robustness and adaptability. By learning higher-level features directly from data, these methods often deliver more accurate and generalizable skeletonization results, even under complex imaging conditions.

First, [19] addresses Style Overfitting in CNN-based Chinese character recognition, where conventional classifiers falter on unfamiliar font styles. To resolve this, the authors propose a Skeleton Kernel implemented as a fixed, narrow rectangular window that extracts fundamental stroke structure through predetermined weights rather than learned ones, thus capturing essential skeletal features. Experimental evaluations confirm that integrating the Skeleton Kernel into a CNN significantly improves generalization performance, even with limited training samples. Next, [20] introduces a fully convolutional network (FCN) for skeletonizing handwritten Chinese characters with complex structures by combining holistically-nested feature extraction, regressive dense upsampling convolution (rDUC), and multi-rate dilated fusion (MDF). Training data are synthesized from online handwriting to circumvent the impracticality of manually labeling skeleton pixels in large offline datasets. The proposed method achieves high F-measure scores in skeleton pixel detection, effectively preserving character shape and visual topology while outperforming traditional thinning algorithms.

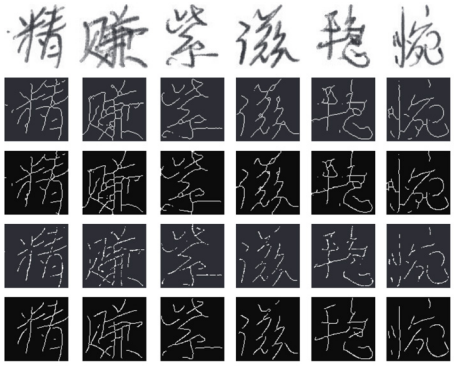
Following this, the paper [21] proposed skeletonization module leverages a lightweight dual FCN architecture to generate both cross and skeleton maps from online handwriting data. It pre-extracts and thins crossing regions using dilated stroke intersections, effectively addressing challenges such as stroke width variation and spurious thinning. A final skeletonization algorithm then refines these maps to accurately preserve the essential structural details of handwritten strokes. These precise skeleton maps provide a reliable basis for subsequent analysis in identifying abnormal strokes within the DeepAD framework. Experimental evaluations on synthesized training data demonstrate that the dual FCNs outperform state-of-the-art models. Overall, the enhanced skeletonization method offers practical and robust improvements, making it a valuable tool for writing quality assessment in educational applications. Additionally, [22] builds on the FCN-based skeletonization models SkeNet and CrossNet [20], [21], presenting a comprehensive stroke extraction scheme for offline handwritten Chinese characters, providing accurate structural information for applications such as education and calligraphy analysis.

Meanwhile, [23] proposes a novel iterative knowledge transfer framework that generates realistic skeleton representations from offline handwriting images. It initially creates synthetic skeletons from offline handwriting (CVL), then employs a reverse mapping network (NN1) trained on these synthetic samples to correct errors, subsequently converting real online skeleton data

(IAM-Online) into synthetic offline samples. A final mapping network (NN2) transforms the offline handwriting images into high-quality skeletons, effectively bridging the gap between offline and online data. Also, [24] presents the JointFontGAN framework, a multi-stream extended conditional GAN (XcGAN) model that learns from limited samples to generate both font skeletons and glyph representations. Its Sk-XcGAN subnetwork predicts generalized skeletons, providing robust structural guidance, while the Con-XcGAN subnetwork enriches these skeletons with detailed glyph content. By integrating the two, the system effectively synthesizes both standard and artistic/handwritten styles while preserving high geometric accuracy and content consistency. [25] introduces RCRN, a GAN-based model that tackles real-world character image degradation and complex noise by integrating character skeleton extraction with scale-ensemble feature analysis. Its architecture consists of a skeleton extractor, SENet, which robustly generates skeletons from degraded images using physics-guided denoising and iterative weighted fusion, and a character image restorer, CiRNet, which reconstructs clean images by combining the original input with the extracted skeletons. Experiments on a newly constructed dataset of 1,606 degraded character images demonstrate that RCRN surpasses state-of-the-art methods both quantitatively and qualitatively.



(a) [23]



(b) Result of [20]. From up to down: input images; stroke continuity based algorithm; ZhangSuen method; HED network, and theirs.

Fig. 2: Some Result From Selected Papers

C. Skeleton applications

Building on these foundations, recent studies show that skeleton-based pre-processing—using simplified images that retain essential structure—can be effectively applied to tasks such as graphology classification, handwriting restoration, and recognition. For instance, [26] represents handwritten Chinese characters as skeleton graphs and applies a Pyramid Graph Transformer combining global attention with local graph convolutions, thereby capturing geometric semantics and achieving superior recognition. In a related study, [27] quantitatively evaluates skeletonization through similarity measures (e.g., correlation coefficient, pixel coincidence, cosine similarity) and then leverages these features within an artificial neural network to enhance recognition. Beyond recognition, skeleton extraction supports handwriting synthesis and calligraphic font generation, as demonstrated by [28], which employs a skeleton-based font-matching strategy, and by SCI-Font [29], which integrates skeleton, contour, and inexact paired data to improve content-style representation. A skeleton guided channel expansion module (SGCE) [30] further addresses mode collapse in GAN-based Chinese font generation. Many of these studies, including [31], [32], employ Zhang-Suen [1] for skeleton preprocessing. Additionally, [33] demonstrates dynamic writing restoration model from static Chinese calligraphy images by extracting an initial skeleton [34], removing distortions, and simulating real pen movements.

IV. Conclusions and Future Works

In this survey, we provided an overview of skeleton extraction methods in handwriting analysis and one of its applied branches-graphology, encompassing both traditional thinning algorithms and modern deep learning approaches. Traditional methods reduce images to single-pixel-wide structures, standardizing handwritten and printed text. However, these techniques often suffer from anisotropy, noise sensitivity, and local distortions. To mitigate these issues, recent algorithms incorporate advanced techniques like ring radius transforms, triangulation, loop analysis, and specialized post-processing to improve accuracy and robustness. Concurrently, deep learning approaches, including CNN-based and GAN-based models, leverage high-level feature learning to adapt to noise, degradation, and style variations, thereby enhancing skeletonization performance. These methods have significantly advanced applications in handwriting restoration, trajectory recovery, and calligraphy evaluation.

Despite these improvements, challenges remain in processing severely degraded images and the dependency on large, high-quality datasets. Future research should focus on exploring unsupervised or semi-supervised learning techniques to address data scarcity and improve scalability across diverse scripts.

Acknowledgment

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] T. Y. Zhang and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Communications of the ACM*, vol. 27, no. 3, pp. 236–239, 1984.
[2] C. Arcelli and G. S. Di Baja, "A width-independent fast thinning algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 4, pp. 463–474, 1985.
[3] M. A. Alghamdi and W. J. Teahan, "A new thinning algorithm for arabic script," *International Journal of Computer Science and Information Security*, vol. 15, no. 1, 2017.
[4] C. Arcelli and G. S. Di Baja, "A one-pass two-operation process to detect the skeletal pixels on the 4-distance transform," *IEEE transactions on pattern analysis and machine intelligence*, vol. 11, no. 4, pp. 411–414, 1989.
[5] J. J. Zou and H. Yan, "Skeletonization of ribbon-like shapes based on regularity and singularity analyses," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 31, no. 3, pp. 401–407, 2001.
[6] S. Singh and N. K. Garg, "Performance analysis of thinning algorithms for offline-handwritten devanagari words," in *International Conference on Advances in Electrical and Computer Technologies*. Springer, 2021, pp. 437–454.
[7] Z. Guo and R. W. Hall, "Parallel thinning with two-subiteration algorithms," *Communications of the ACM*, vol. 32, no. 3, pp. 359–373, 1989.
[8] T.-C. Lee, R. L. Kashyap, and C.-N. Chu, "Building skeleton models via 3-d medial surface axis thinning algorithms," *CVGIP: graphical models and image processing*, vol. 56, no. 6, pp. 462–478, 1994.
[9] R. Plamondon, C. Y. Suen, M. Bourdeau, and C. Barrière, "Methodologies for evaluating thinning algorithms for character recognition," *International journal of pattern recognition and artificial intelligence*, vol. 7, no. 05, pp. 1247–1270, 1993.
[10] P. K. Saha, G. Borgefors, and G. S. di Baja, "A survey on skeletonization algorithms and their applications," *Pattern recognition letters*, vol. 76, pp. 3–12, 2016.
[11] V. Pervouchine, G. Leedham, and K. Melikhov, "Handwritten character skeletonisation for forensic document analysis," in *Proceedings of the 2005 ACM symposium on Applied computing*, 2005, pp. 754–758.
[12] X. Zhang, M. Wang, L. Wang, Q. Huo, and H. Li, "Building handwriting recognizers by leveraging skeletons of both offline and online samples," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2015, pp. 406–410.
[13] K. Fujita, "Extract an essential skeleton of a character as a graph from a character image," *arXiv preprint* arXiv:1506.05068, 2015.
[14] V. A. Kha, H. H. Kha, and M. Blumenstein, "Extraction of dynamic trajectory on multi-stroke static handwriting images using loop analysis and skeletal graph model," *REV Journal on Electronics and Communications*, vol. 6, no. 1-2, 2016.
[15] Y. Wu, P. Shivakumara, W. Wei, T. Lu, and U. Pal, "A new ring radius transform-based thinning method for multi-oriented video characters," *International journal on document analysis and recognition (IJDAR)*, vol. 18, pp. 137–151, 2015.
[16] D. Phan, I.-S. Na, S.-H. Kim, G.-S. Lee, and H.-J. Yang, "Triangulation based skeletonization and trajectory recovery for handwritten character patterns," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 9, no. 1, pp. 358–377, 2015.
[17] M. Dinh, H.-J. Yang, G.-S. Lee, S.-H. Kim, and L.-N. Do, "Recovery of drawing order from multi-stroke english handwritten images based on graph models and ambiguous zone analysis," *Expert Systems with Applications*, vol. 64, pp. 352–364, 2016.
[18] Z. Xu, F. Qin, N. Xiong, H. Shi, J. Zhang, and W. Lin, "Skeleton extraction of hard-pen regular script based on stroke characterization and ambiguous zone detection," *IET Image Processing*, vol. 18, no. 6, pp. 1504–1515, 2024.
[19] W. Tang, Y. Su, X. Li, D. Zha, W. Jiang, N. Gao, and J. Xiang, "Cnn-based chinese character recognition with skeleton feature," in *Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part V 25*. Springer, 2018, pp. 461–472.
[20] T.-Q. Wang and C.-L. Liu, "Fully convolutional network based skeletonization for handwritten chinese characters," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.
[21] T.-Q. Wang and C. L. Liu, "Deepad: A deep learning based approach to stroke-level abnormality detection in handwritten chinese character recognition," in *2018 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2018, pp. 1302–1307.
[22] T.-Q. Wang, X. Jiang, and C.-L. Liu, "Query pixel guided stroke extraction with model-based matching for offline handwritten chinese characters," *Pattern Recognition*, vol. 123, p. 108416, 2022.
[23] M. Mayr, M. Stumpf, A. Nicolaou, M. Seuret, A. Maier, and V. Christlein, "Spatio-temporal handwriting imitation," in *European Conference on Computer Vision*. Springer, 2020, pp. 528–543.
[24] Y. Xi, G. Yan, J. Hua, and Z. Zhong, "Jointfontgan: Joint geometry-content gan for font generation via few-shot learning," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 4309–4317.
[25] D. Shi, X. Diao, H. Tang, X. Li, H. Xing, and H. Xu, "Rcrn: Real-world character image restoration network via skeleton extraction," in *Proceedings of the 30th ACM international conference on multimedia*, 2022, pp. 1177–1185.
[26] J. Gan, Y. Chen, B. Hu, J. Leng, W. Wang, and X. Gao, "Characters as graphs: Interpretable handwritten chinese character recognition via pyramid graph transformer," *Pattern Recognition*, vol. 137, p. 109317, 2023.
[27] Y. Zhao, X. Zhang, B. Fu, Z. Zhan, H. Sun, L. Li, and G. Zhang, "Evaluation and recognition of handwritten chinese characters based on similarities," *Applied Sciences*, vol. 12, no. 17, p. 8521, 2022.
[28] D. G. Balreira and M. Walter, "Handwriting synthesis from public fonts," in *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2017, pp. 246–253.
[29] Y. Zhang, Y. Wang, J. Xiong, J. Zhou, and J. Zeng, "Sci-font: Enhancing content-style representation for chinese calligraphy generation with skeleton, contour and inexact paired data," in *International Conference on Artificial Neural Networks*. Springer, 2024, pp. 377–391.
[30] J. Zhou, Y. Wang, Y. Yuan, Q. Huang, and J. Zeng, "SGCE-font: Skeleton guided channel expansion for chinese font generation," *arXiv preprint* arXiv:2211.14475, 2022.
[31] W. Chen, X. Su, and H. Zhang, "Script-level word sample augmentation for few-shot handwritten text recognition," in *International Conference on Frontiers in Handwriting Recognition*. Springer, 2022, pp. 316–330.
[32] C. Chan, "Stroke extraction for offline handwritten mathematical expression recognition," *IEEE Access*, vol. 8, pp. 61 565–61 575, 2020.
[33] W. Zhang, W. An, and D. Huang, "Writing process restoration of chinese calligraphy images based on skeleton extraction," in *IOP Conference Series: Materials Science and Engineering*, vol. 569, no. 3. IOP Publishing, 2019, p. 032060.
[34] R. Stefanelli and A. Rosenfeld, "Some parallel thinning algorithms for digital pictures," *Journal of the ACM (JACM)*, vol. 18, no. 2, pp. 255–264, 1971.

Age Estimation from Handwriting Images with Regression Models

Lior Abergel, Maor Merling, Marina Litvak, Irina Rabaev
Software Engineering Department, Shamoon College of Engineering
Beer Sheva, 84100, Israel
{liorab3,maorme,marinal,irinar}@ac.sce.ac.il

Abstract

Accurately estimating a person's age from handwriting presents a unique challenge in the fields of biometrics, forensic science, and cognitive assessment. Traditional approaches rely on classification models that assign writers to predefined age groups, limiting precision and flexibility. In this study, we introduce a regression-based deep learning framework for continuous age prediction, leveraging state-of-the-art convolutional neural networks (CNNs). Our approach utilizes ResNet50, InceptionV3, and DenseNet121, along with their ensemble, to enhance predictive performance. Experiments on the HHD dataset demonstrate that ResNet50 outperforms other CNNs in seven out of nine metrics. Furthermore, the ResNet50 and DensNet121 ensemble outperforms other architectures and ensembles in five of ten metrics, achieving the lowest Mean Absolute Error (MAE) of 2.93 years and setting a new benchmark for handwriting-based age estimation.

Index Terms

Handwriting analysis, Age estimation, Regression, Ensemble of Models, ResNet50, InceptionV3, DenseNet121

I. Introduction

Handwriting analysis has long been a tool for studying demographic characteristics, with applications spanning forensic investigations [16], medical diagnostics [4], [10], psychology and education [7], [8], and biometrics [6], [17]. This paper presents a study aimed at estimating age based on handwriting using advanced deep-learning methodologies.

In previous research, handwriting-based age estimation has been treated as a classification problem in which age is divided into discrete groups (usually from two to four), and a model predicts the most likely group. This classification approach is motivated by findings in graphology and developmental psychology, which show that handwriting undergoes observable transitions at different life stages—such as the maturation of writing in adolescence and its decline in older age. For example, AL-Qawasmeh et al. [2] justify categorizing writers as "youth adults" and "mature adults" based on graphologist-identified features such as slant and pressure, especially in forensic and clinical contexts where detecting abnormal vs. normative aging patterns is more relevant than pinpointing exact age. Similarly, Basavaraja et al. [5] divide writers into four adolescent and young adult groups, linking disconnectedness in handwriting to cognitive load and psychological stress as age increases. Asci et al. [4] also support this view, showing that stroke size and handwriting regularity degrade with age due to motor and neurological decline, warranting broad age-based groupings. Moreover, practical considerations such as limited labeled data and forensic application needs have reinforced the use of discrete categories.

However, because age is a continuous variable, grouping can obscure fine-grained differences between individuals. We hypothesize that regression-based methods are better suited for age estimation task. Moreover, regression has been successfully employed for age estimation in other domains. For example, AL-Qawasmeh et al. [9] applied regression for facial image analysis to detect both age and gender. Asci et al. [4] used logistic regression for handwriting-based detection of Parkinson's disease—underscoring the versatility of regression in similar tasks. In neurological studies focusing on brain age prediction, various regression models have consistently shown strong performance, including Lasso regression [11], [14], Support Vector Regression (SVR) [11], [12], Relevance Vector Regression (RVR) [11], [12], and Gaussian Process Regression (GPR) [12]. Integrated learning methods, featuring combined classification and regression have further demonstrated predictive robustness in age-related tasks [12], [15].

To the best of our knowledge, the current study is the first to approach age estimation from handwriting as a regression problem. Beyond methodological novelty, exact age estimation has concrete, high-impact uses. Precise writer's age estimation of anonymous or disputed handwritten material can help narrow a pool of suspects. It can also assist document-authentication experts in establishing provenance for historical archives. In clinical settings, the divergence between chronological and handwriting-predicted age may act as a marker of accelerated neuro degeneration, while in education, it can help tailor developmental interventions for children. These scenarios all demand resolution finer than broad age classes, justifying our regression-based approach.

II. Related Work

The challenge of age estimation from handwriting has been approached using various computational models, from traditional machine learning techniques to modern deep learning architectures.This section reviews the relevant literature, with key details summarized in Table I.

TABLE I
SUMMARY OF HANDWRITING-BASED AGE CLASSIFICATION STUDIES

Authors (Year)	Model / Approach	Dataset	# Age Categories	Acc. (%)
Singla et al. [1]	Inception V3	Gurmukhi script (74 760 samples)	Three age groups: 5–10, 10–15, 15–20	92.7
AL-Qawasmeh et al. [2]	Graphologist features → SVM/NN	FSHS, KHATT	Two age groups: young and mature adults	71.0 (FSHS), 65.2 (KHATT)
Huang et al. [3]	VGG-16 + Harmonic WT features	IAM-2, KHATT, and private dataset	Two age groups for IAM-2 and KHATT, 11 groups 11 five-year spans, 11–65 yrs for the private dataset	71.43 (IAM-2), 64.44 (KHATT), 82.38 (private dataset)
Asci et al. [4]	ResNet-50	private dataset, 156 images (18–90 yrs)	Three age groups: 8–35, 36–55, 56–90	79.0
Basavaraja et al. [5]	Disconnectedness + Iterative K-means + Hu Moments	400 samples	Four age groups: 11–12, 13–16, 17–20, 21–24	67.5
Rabaev et al. [6]	B-ResNet	KHATT (2 000 images)	Four age groups: $\leq$15, 16–25, 26–50, $>$50	66.65
AL-Qawasmeh et al. [7]	ResNet + GoogleNet features → SVM	2 428 Arabic handwriting samples	Two age groups: 16–24, 25–55	69.70
Shin et al. [8]	SVM + Random Forest	138 Japanese handwriting images, 57 texts and 81 patterns (zigzag/periodic-line images)	Two age groups: 8–13, 19–59	93.5 (text), 89.8 (pattern)

Singla et al. [1] utilized the Inception V3 model to classify age based on handwritten characters in the Gurmukhi script. Their dataset of 74,760 handwritten character samples was split into three age groups (5-10, 10-15, 15-20), and the model achieved an accuracy of 92.7% on this classification task. AL-Qawasmeh et al. [2] extracted a set of age-related features, identified by a graphologist, to classify handwritten documents into two age groups: young adults and mature adults. The experiments were conducted with Support Vector Machines (SVM) and Neural Networks (NN) on the FSHS and KHATT datasets, and the reported accuracies are 71% and 65.2%, respectively. Huang et al. [3] trained a VGG-16 model using both row images and features derived from the Harmonic Wavelet Transform. They evaluated performance on four different datasets: IAM-2 and KHATT for two-class classification, a dataset from [5] divided into four age classes, and their own dataset divided into eleven narrower five-year spans classes ranging from 11 to 65. Reported classification accuracies for the two-class setup were 71.43% on IAM-2 and 64.44% on KHATT, and 82.38% on their custom dataset for the 11-class task. Asci et al. [4] applied ResNet-50 to identify and analyze handwriting changes related to age, achieving an accuracy of 79%. The study was conducted on a small dataset of 156 handwriting images covering ages 18–90. Their data was divided into three age groups: young (18-35), middle-aged (36-55), and elderly (56-90). V. Basavaraja et al. [5] combined Disconnectedness Features with Iterative K-means and Hu Moments to cluster 400 handwriting samples into four age groups (11–12, 13–16, 17–20, 21–24), achieving an accuracy rate of 67.5%. Rabaev et al. [6] introduced the B-ResNet model for handwriting-based age and gender classification. Evaluated on the KHATT dataset, containing over 2,000 Arabic and English handwriting images divided into four age groups ($\leq$ 15, 16–25, 26–50, and $>$ 50), their method achieved an accuracy of 66.65%. AL-Qawasmeh et al. [7] combined ResNet and GoogleNet (Inception V1) for feature extraction, using SVM as the classifier. Their approach was tested on 2,428 Arabic handwriting samples split into two age groups (16–24 and 25–55), yielding a classification accuracy of 69.70%. Shin et al. [8] employed a combined SVM and Random Forest (RF) strategy for classifying people as adults or children based on their handwritten text and patterns (zigzag and periodic lines). The reported accuracies are 93.5% and 89.8% for text and pattern classification, respectively. They experimented with a dataset of 138 images (57 handwriting in Japanese and 81 pattern samples) divided into two groups: 8–13 and 19–59.

As can be noted, the methods overviewed above address age estimation from handwriting as a classification task, typically restricting the problem to 2–4 age classes.

III. Methodology

In this study, we experimented with three state-of-the-art deep-learning models—ResNet50 [19], InceptionV3 [20], and DenseNet121 [21]–for the task of age estimation from handwriting. These models were chosen because prior research has

demonstrated their effectiveness in extracting shape- and structure-related features. ResNet50 and DenseNet121 have been widely applied to tasks requiring the identification of fine-grained features. InceptionV3 excels at capturing multi-scale features. Additionally, we used ensembles that combine all three or only two of them, leveraging complementary feature representations to enhance overall accuracy. We follow the constant weighting approach [24], where the final output (predicted age) is calculated from the individual outputs (age predicted by the individual models). We apply arithmetic and a weighted average as integration functions, where weak predictors have less influence on the final prediction.

A. Dataset Preparation

We utilized the HHD dataset [18], which contains handwritten samples in Hebrew together with the age labels. The HHD dataset comprises approximately 880 handwriting samples from writers aged 8 to 60, with the majority being high school students. Figure 1 presents the distribution of ages in the HHD dataset.

The dataset was split into training, test, and validation sets using the $70 : 15 : 15$ ratio.

B. Experimental settings

We utilized the transfer-learning approach. All three models utilized in this study were pre-trained on ImageNet. To adapt the models for the age regression task, the fully connected classification layer was removed and replaced by a global average pooling layer, followed by a dropout layer with 0.5 ratio, and the output layer with a single neuron. A linear activation function was used, ensuring the model outputs a continuous age value.

For all the experiments, we used a learning rate of $1e-3$ and a batch size of 128. Initially, the pre-trained weights were frozen, and the models were trained for 50 epochs. Then, the weights were unfrozen, and the models were trained again for an additional 10 epochs.

Data preprocessing involved resizing the images to a height of 800 pixels while maintaining their aspect ratio and normalizing pixel values to the [0,1] range. Each image was divided into overlapping patches of 400×400 pixels using a stride of 200 pixels. To enhance model generalization, we applied various data augmentation techniques during training, including rotation, zoom, brightness adjustment, and Gaussian noise.

C. Evaluation Metrics

The performance of the models was assessed using several metrics: MAE, $RMSE$, R^2 score, $MAPE$ (%), threshold-based accuracy (the percentage of correct predictions within 2, 5, and 10 years of the true age), and additional statistics such as the maximum and median errors.

Metrics such as MAE and $RMSE$ (Eq. 1, 2 below) are particularly suited for this task because they provide a direct and intuitive measure of how close the predicted ages are to the true values. Both range from $[0, \infty)$, with lower values indicating better model performance. MAE measures the average magnitude of errors, while RMSE gives a higher weight to large errors, making it more sensitive to outliers. The coefficient of determination R^2 (Eq. 3) helps to evaluate the quality of the model's features and its ability to explain the variance in the target variable. It ranges from $(-\infty, 1]$, with $R^2 = 1$ indicating a perfect fit, meaning the model explains all the variance in the target variable, and $R^2 = 0$ suggesting that the model performs no better than the mean of the target values. Negative values indicate that the model performs worse than simply predicting the mean.

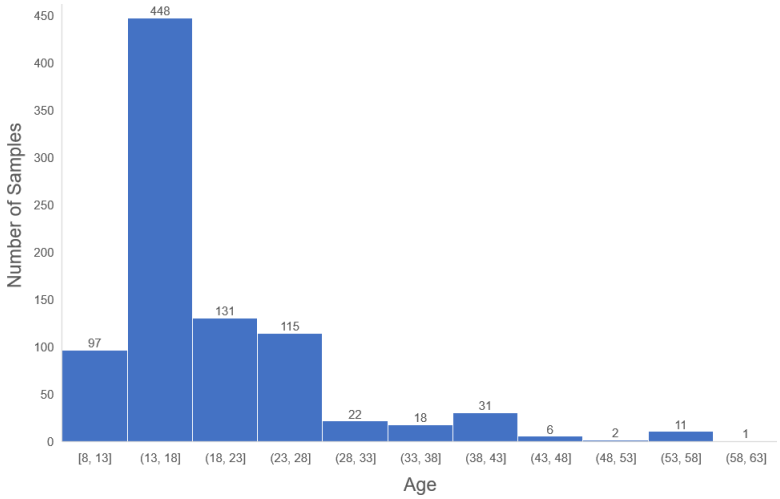


Fig. 1. Age distribution of the HHD dataset.

Values smaller than -1 arise whenever the residual sum of squares is more than twice the total sum of squares, implying a very poor model [25]. Relative metrics such as $MAPE$ (Eq. 4) provide insight into the precision of predictions across different age groups, especially when there is a significant variance between them. $MAPE$ measures the average percentage error between the predicted and actual values and is particularly useful when the target variable has a wide range. It ranges from $[0, \infty)$, expressed as a percentage. Lower $MAPE$ values are better, indicating more accurate and reliable predictions.

These metrics collectively provide a comprehensive evaluation of model performance, highlighting both the accuracy and robustness of predictions across different age groups.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (3)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \quad (4)$$

Here, y_i denotes the actual values, $\hat{y}_i$ the predicted values, $\bar{y}$ the mean of the actual values, and n is the total number of samples. The indicator function $\mathbb{I}(\cdot)$ returns 1 if the condition inside is satisfied, otherwise 0.

IV. Results and Discussion

Table II contains the comparative results of the three chosen models and their ensembles. We tried the full ensemble that integrates all three models and partial ensembles that integrate only two best-performing predictors – ResNet and DenseNet. Both ResNet+Inception+DensNet and ResNet+DensNet apply equal weights for each model. The weights assigned to each predictor in the weighted average were set up empirically. The best scores among individual models are colored in grey, and the best scores among all six methods (including ensembles) are marked in bold.

Model	MAE	RMSE	R2	MAPE	acc±2	acc±5	acc±10	Max Err	Med Err
ResNet50	3.23	6.03	0.09	18.13	56.9	81.9	**96.55**	30.6	1.59
InceptionV3	5.06	6.55	-0.07	31.29	6.9	67.24	93.97	27.78	4.12
DenseNet121	3.97	6.1	0.07	22.8	33.62	**82.76**	93.1	**26.85**	2.76
ResNet+Inception+DensNet	3.1	**5.81**	0.16	17.78	57.76	81.9	**96.55**	28.41	1.33
ResNet+DensNet	3.08	**5.81**	**0.15**	16.99	65.52	81.03	**96.55**	28.72	1.44
Resnet+DensNet_$weigthed$	**2.93**	5.84	**0.15**	**16.12**	**66.38**	81.03	**96.55**	29.48	**1.3**

TABLE II
PERFORMANCE METRICS OF THE MODELS

As can be seen in the table, ResNet50 outperforms other individual models in most (7 out of 10) of the metrics, while InceptionV3 has the worst performance out of the three models. Both ResNet50 and DenseNet121 show low MAE, RMSE, MAPE, median error, positive R^2, and high within ±2 / ±5 years accuracy and cumulative error, proving their ability to predict document age more precisely than InceptionV3. ResNet50 and DenseNet121 are superior in age prediction from scanned handwriting due to their strong feature extraction and efficient learning mechanisms. ResNet50's deep residual learning captures fine-grained handwriting details, while DenseNet121's dense connections enhance feature propagation, making both models highly effective in recognizing handwriting degradation patterns. The weighted ensemble of these two predictors (WRD, with weights 0.7 and 0.3 for ResNet and DenseNet, respectively) achieves the best performance with the lowest MAE (2.93) and highest Within ±2 Years accuracy (66.38%), demonstrating that combining these architectures further improves the accuracy of the prediction.

Our results align with trends reported in prior category-based age classification studies. For example, Asci et al. [4] demonstrated that a CNN using a ResNet-50 backbone achieved up to 79% accuracy when distinguishing young, middle-aged, and older adults, confirming the effectiveness of this architecture for handwriting-driven age inference.

Error analysis revealed that all the models struggled to predict age for an age group "> 26". Two factors can plausibly explain this. (1) Dataset imbalance: the long-tail distribution of subjects aged over 26 (see Figure 1), resulting in much fewer observations within this age band. With a limited number of examples, the regressor struggles to learn the subtle cues of mid and late adulthood. (2) Developmental plateau. Handwriting kinematics stabilize by early adulthood, showing minimal change until late-life motor decline becomes detectable [4], [26]–[28]. The reduced changes in motor control signal in the 26–50 window make that age band inherently harder to predict. This combination of low data representation and minimal age-related biomechanical signal likely drives the model's errors for the "> 26" age group.

V. Conclusion

This study explored the application of deep learning regression models for age estimation from handwriting, demonstrating that ResNet50 and DenseNet121 significantly outperform other architectures. By leveraging their strong feature extraction

capabilities, these models effectively captured handwriting degradation patterns. The experimental results confirmed that regression-based approaches, particularly ensemble methods, provide more precise age predictions compared to traditional discrete age group classification. The weighted ensemble of ResNet and DensNet achieved the best performance, with the lowest MAE (2.93) and the highest accuracy within ±2 years (66.38%). These findings highlight the potential of deep learning in forensic analysis, cognitive research, and historical document studies. Future work can explore larger and more diverse datasets, incorporate additional handwriting features, and refine ensemble techniques to further enhance predictive accuracy. The insights gained from this research contribute to academic understanding and practical applications in fields relying on handwriting analysis for identity verification, cognitive assessments, and developmental studies.

References

[1] C. Singla, R. Maini, and M. Kumar, "Age Classification through Handwriting Analysis using Convolutional Neural Networks and Pre-Segmented Offline Handwritten Gurumukhi Characters," in Proc. Seventh Int. Conf. Image Inf. Process. (ICIIP) , 2023, pp. 51–58. doi: 10.1109/ICIIP61524.2023.10537631.

[2] N. AL-Qawasmeh, M. Khayyat, and C. Y. Suen, "Age Detection From Handwriting Using Different Feature Classification Models," Pattern Recognit. Lett. , vol. 167, pp. 60–66, 2023. doi: 10.1016/j.patrec.2023.02.001.

[3] Z. Huang, P. Shivakumara, M. A. Kaljahi, A. Kumar, U. Pal, and T. Lu, "Writer Age Estimation Through Handwriting," Multimedia Tools Appl. , vol. 82, pp. 16033–16055, 2023. doi: 10.1007/s11042-022-13840-w.

[4] F. Asci, S. Scardapane, A. Zampogna, V. D'Onofrio, L. Testa, M. Patera, M. Falletti, L. Marsili, and A. Suppa, "Handwriting Declines With Human Aging: A Machine Learning Study," Front. Aging Neurosci. , vol. 14, p. 889930, 2022. doi: 10.3389/fnagi.2022.889930.

[5] V. Basavaraja, P. Shivakumara, D. S. Guru, U. Pal, and M. Blumenstein, "Age Estimation Using Disconnectedness Features in Handwriting," in Proc. Int. Conf. Document Anal. Recognit. (ICDAR) , 2019, pp. 1130–1135. doi: 10.1109/ICDAR.2019.00183.

[6] I. Rabaev, I. Alkoran, O. Wattad, and M. Litvak, "Automatic Gender and Age Classification from Offline Handwriting with Bilinear ResNet," Sensors , vol. 22, p. 9650, 2022. doi: 10.3390/s22249650.

[7] N. AL-Qawasmeh and C. Y. Suen, "Transfer Learning to Detect Age From Handwriting," Adv. Artif. Intell. Mach. Learn., vol.2, no.2, pp. 394–406, 2022.

[8] J. Shin, M. Maniruzzaman, Y. Uchida, et al., "Important Features Selection and Classification of Adult and Child from Handwriting Using Machine Learning Methods," Appl. Sci., vol. 12, no. 5256, 2022. doi: 10.3390/app12105256.

[9] K. Ito, H. Kawai, T. Okano, and T. Aoki, "Age and Gender Prediction from Face Images Using Convolutional Neural Network," in Proc. APSIPA Annu. Summit Conf. 2018 , 2018, pp. 7–11.doi: 10.23919/APSIPA.2018.8659655.

[10] C. Poon, N. Gorji, M. Latt, K. Tsoi, S. B. Choy, C. Loy, and S. Poon, "Derivation and Analysis of Dynamic Handwriting Features as Clinical Markers of Parkinson's Disease," in Proc. 52nd Hawaii Int. Conf. Syst. Sci. , 2019, pp. 3721–3730.

[11] M. Xiong, L. Lin, Y. Jin, W. Kang, S. Wu, and S. Sun, "Comparison of Machine Learning Models for Brain Age Prediction Using Six Imaging Modalities on Middle-Aged and Older Adults," Sensors , vol. 23, no. 3622, 2023. doi: 10.3390/s23073622.

[12] M. S. Treder, J. P. Shock, D. J. Stein, S. du Plessis, S. Seedat, and K. A. Tsvetanov, "Correlation Constraints for Regression Models: Controlling Bias in Brain Age Prediction," Front. Psychiatry , vol. 12, p. 615754, 2021. doi: 10.3389/fpsyt.2021.615754.

[13] F. Alaei and A. Alaei, "Review of Age and Gender Detection Methods Based on Handwriting Analysis," Neural Comput. Appl. , vol. 35, pp. 23909–23925, 2023. doi: 10.1007/s00521-023-08996-x.

[14] L. Baecker, J. Dafflon, P. F. da Costa, R. Garcia-Dias, S. Vieira, C. Scarpazza, and W. H. L. Pinaya, "Brain age prediction: A comparison between machine learning models using region- and voxel-based morphometric data," Hum. Brain Mapp. , vol. 42, no. 9, pp. 2332–2346, 2021. doi: 10.1002/hbm.25368.

[15] J. Chen, L. Cheng, X. Yang, J. Liang, B. Quan, and S. Li, "Joint Learning with Both Classification and Regression Models for Age Prediction," J. Phys.: Conf. Ser. , vol. 1168, no. 3, p. 032016, 2019. doi: 10.1088/1742-6596/1168/3/032016.

[16] A. U. Anupama and D. Mahna, "Handwriting Examination in Forensic Science," Pak. Heart J. , vol. 56, no. 3, pp. 1298–1302, 2023. Available: https://www.pkheartjournal.com/index.php/journal/article/view/1837.

[17] M. Faundez-Zanuy, J. Fierrez, M. A. Ferrer, M. Diaz, R. Tolosana, and R. Plamondon, "Handwriting Biometrics: Applications and Future Trends in e-Security and e-Health," Cogn. Comput. , vol. 12, pp. 940–953, 2020. doi: 10.1007/s12559-020-09755-z.

[18] I. Rabaev, B. K. Barakat, A. Churkin, and J. El-Sana, "The HHD Dataset," in Proc. 16th Int. Conf. Front. Handwriting Recognit. (ICFHR) , 2020. doi: 10.1109/ICFHR2020.2020.00050.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) , 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) , 2016, pp. 2818–2826. doi: 10.1109/CVPR.2016.308.

[21] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) , 2017, pp. 4700–4708. doi: 10.1109/CVPR.2017.243.

[22] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," in Proc. AAAI Conf. Artif. Intell. (AAAI) , 2017, pp. 4278–4284.

[23] M. Tan and Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," in Proc. Int. Conf. Mach. Learn. (ICML) , 2021, pp. 10096–10106. doi: 10.48550/arXiv.2104.00298.

[24] J. Mendes-Moreira, C. Soares, A. M. Jorge, and J. F. De Sousa, "Ensemble approaches for regression: A survey," in ACM Computing Surveys (CSUR), vol. 45, no. 1, article no. 10, pp. 1–40, Dec. 2012. doi: 10.1145/2379776.2379786.

[25] T. O. Kvålseth, "Cautionary Note about R^2," *Am. Stat.*, vol. 39, no. 4, pp. 279–285. https://doi.org/10.1080/00031305.1985.10479448

[26] S. Graham, V. Berninger, N. Weintraub, and W.Schafer, "Development of handwriting speed and legibility in Grades 1–9," *Journal of Educational Research*, vol.,92, no.,1, pp.,42–52, 1998. doi: 10.1080/00220679809597574.

[27] Zhang, X., Huang, M., Yuan, X., Zhong, X., Dai, S., Wang, Y., Zhang, Q., Wongwitwichote, K., & Jiang, C. (2025). Lifespan trajectories of motor control and neural oscillations: A systematic review of magnetoencephalography insights. *Developmental Cognitive Neuroscience*. doi: 10.1016/j.dcn.2025.101529

[28] Plamondon, R., O'Reilly, C., Rémi, C., & Duval, T. (2013). The lognormal handwriter: Learning, performing, and declining. *Frontiers in Psychology*, *4*, 945. doi: 10.3389/fpsyg.2013.00945

Benchmarking Facial Landmarks Estimation: Evaluating Popular Algorithms Using FAIRSET, a Balanced Landmark Dataset

Ian-Mathieu Joly[1,3], Nikola Zelovic[1,3], Eléonor Riesco[2,3], Karina Lebel[* 1,3]

[1]Faculty of Engineering, Université de Sherbrooke, Sherbrooke, QC, Canada, J1K 2R1
[2]Faculty of Physical Activity, Université de Sherbrooke, Sherbrooke, QC, Canada, J1K 2R1
[3]Research Center on Aging, CIUSS de l'Estrie - CHUS, Sherbrooke, QC, Canada, J1H 4C4

[*]karina.lebel@usherbrooke.ca

***Abstract*— Deep learning-based face landmarking algorithms are widely used in movement analysis, but their accuracy varies according to different factors. This study uses FAIRSET to evaluate the robustness of popular algorithms, identifying and confirming demographic biases through statistical analysis. Based on these findings, we propose guidelines for selecting a lgorithms that account for said biases.**

Keywords*— *Face Landmarking, Facial Analysis, Deep Learning bias, Inclusive Dataset, AI Algorithms Evaluation.

I. Introduction

In recent years, deep learning-based face landmarking algorithms have found their use in multiple movement analysis problems, ranging from entertainment purposes like animation to more serious contexts such as medical studies [1-2]. Their use in these studies has highlighted the need for high accuracy in complex tasks. However, the accuracy of these algorithms can vary significantly depending on factors related to the environment in which the video's images were captured, like occlusions or lighting, but also based on demographics such as skin tone and sex [3-4]. This variation partially comes from imbalances in the training data, as inclusivity is often not at the forefront when developing datasets. Many datasets exist for face landmarking tasks, but the most standard is 300W, which combined and annotated the preexisting AFW, LFPW, HELEN, and XM2VTS datasets [5]. It was the first dataset to include the 68 keypoints seen in most algorithms and comprises "real-world images" to mitigate the effects of challenging environments. Other datasets were created to address specific issues, such as COFW for occlusions or Menpo for head poses. While some datasets include age or skin tone labels, none were specifically designed to address demographic issues. For other computer vision tasks such as face detection (without lankmarking) or sex classification, datasets like UTKFace or FotW aimed to include a wide variety of ages, sexes, and ethnicities, successfully reducing performance bias based on these factors [6-7]. Benchmarks such as the 300W challenge do not include demographic factor labels, focusing instead on environmental factors [8]. Consequently, algorithms generally do not disclose their performances across different demographics, making them difficult to use in contexts where these factors matter. In the face of this issue, the FAIRSET dataset (Face Annotations, Inclusivity & Representation dataSET) was designed as a benchmark to evaluate different algorithms based on demographics. It includes age, sex, and skin tone labels on images annotated with 36 different landmarks, 31 of which are on the face [9]. This dataset enables users to assess whether an algorithm's results vary based on demographics and identify demographic groups where performance declines, thus affecting usability.

Following this work, the present paper uses FAIRSET to evaluate the robustness of popular face landmarking algorithms across different demographics. Specifically, (1) Validating whether a person's demographics affects the accuracy of popular face landmarking algorithms; and (2) Proposing guidance for selecting appropriate algorithm based on project characteristics and validation results using FAIRSET.

II. Methodology

Algorithm choices were based on their popularity rather than solely on state-of-the-art (SOTA) performance. Using popular algorithms is more appropriate since the results will affect more current applications, and these algorithms tend to remain relevant longer. New algorithms that outperform the previous generation are constantly emerging, making SOTA algorithms more fleeting compared to easier-to-use, more popular algorithms. The algorithms selected for this study are Google's Medipapipe [10], 3DDFAv3 [11], and the older OpenFace2.2 [12]. The normalized mean error (NME) was used to evaluate the accuracy of these algorithms. This metric is defined as the distance between a landmark and its ground truth normalized with the distance between the outer corners of a person's eyes [5].

Mediapipe's FaceMeshV2 algorithm returns a mesh of 478 3D points in pixel coordinates, with a reported NME of 3.11% on their proprietary dataset [10]. The latest model available on Google's website is evaluated here. Notably, Google provides a

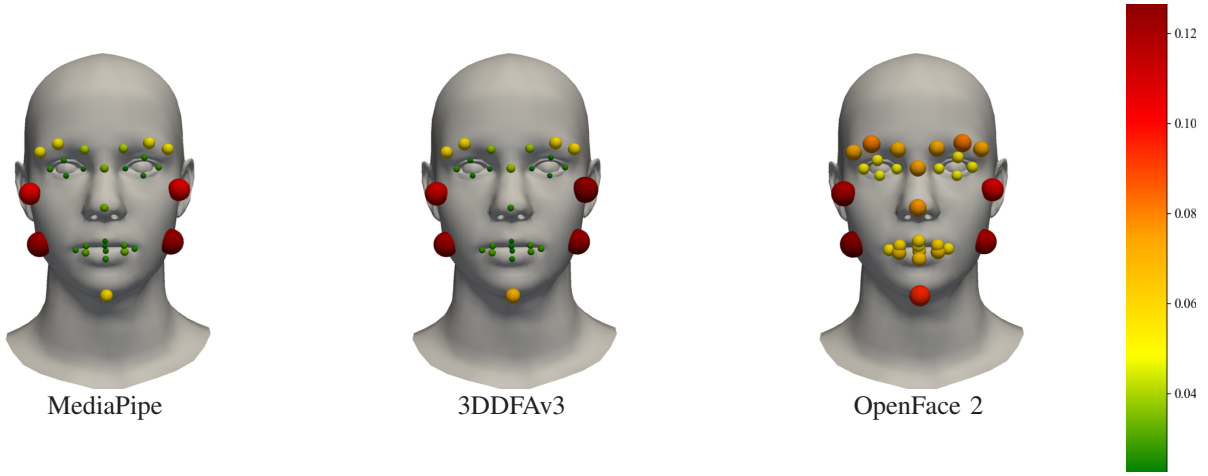


Fig. 1. Keypoint Normalized Mean Error by Algorithm

mean error by skin tone on their dataset, giving some grounds for comparison with our results. 3DDFA-V3 is the latest in the 3DDFA series of algorithms. Its release paper does not provide an NME metric, focusing instead on a 3D error in millimeters for specific face regions. It reports an average error of 1.436 mm for a person's front view and 1.442 mm for side views per region, based on a proprietary dataset [11]. OpenFace2 was trained on the Helen and LFPW datasets, which are part of the 300W dataset. It provides multiple error plots, though only one for the full 68 keypoints showing a median mean squared error (MSE) of roughly 0.04 [12]. This evaluation was conducted on a more obscure dataset named IJB-FL, which focuses on challenging head poses [13].

Landmarks were extracted similarly across the algorithms by mapping the keypoints of each algorithm to the dataset. MediaPipe had an advantage due to its additional keypoints, which allowed for a mapping selection that better aligned with FAIRSET's points than the 68 keypoints algorithms. To avoid statistical biases, the NME offset was removed by keypoint from subsequent results. For images with multiple people, an association was made based on the intersection over union (IOU) of the FAIRSET and the detected bounding box. It is important to note that no confidence metrics or other quality metrics, such as MediaPipe's visibility, were used to filter out estimations, as setting a threshold is algorithm-dependent and subjective to the user. Finally, images where the algorithms found no detections or estimations were withdrawn from the analysis. Images that failed the association were treated similarly.

To ensure the quality of the results, a preprocessing phase was implemented. Low-resolution images were filtered out to prevent low inter-occular distances (IOD) from affecting the results. Individuals with no IOD or an IOD of less than $50px$ were excluded (n=47). In addition, any keypoint results with an NME higher than 1 were filtered out to remove extreme outliers, such as incorrect associations between the individuals in the dataset and the algorithm's detections (n=123).

III. Results

The total number of images with failed detections and associations for each algorithm was 118 for MediaPipe, 3 for 3DDFAv3, and 46 for OpenFace2, resulting in 135 unique images being removed from further analysis. The overall accuracy of each algorithm is shown in Figure 1, where a keypoint's error is represented by its color and scaled size. This indicates that while OpenFace2 performs significantly worse than the others with an NME of 0.070, MediaPipe and 3DDFAv3 exhibit nearly identical performance with NMEs of 0.044 and 0.045, respectively. Additionally, keypoints along the jawline perform much worse than the rest, with average NMEs of 0.114, 0.120, and 0.119 for MediaPipe, 3DDFav3, and Openface2, respectively. Analyses were performed on three demographic factors: age, sex, and skin tone. The NME for each subgroup was extracted per algorithm. These results are presented in figures 2, 3, and 4 where the mean and median lines are dashed and full, respectively.

To validate the statistical link between NME and demographic factors, a one-way analysis of variance (ANOVA) was selected for factors with more than two groups and an independent t-test was used for the rest. Since ANOVA and t-tests assume homogeneity of variances, and combining normalized keypoints disrupts this assumption, the analyses were done by keypoints for each algorithm. The results are summarized in Table I, which shows the number of keypoints where a statistically significant link was found in the ANOVA or t-test ($p < 0.05$). In other words, it reports how many of FAIRSET's 31 keypoints appear to be influenced by each factor. Environmental factors were also included to provide grounds for comparison.

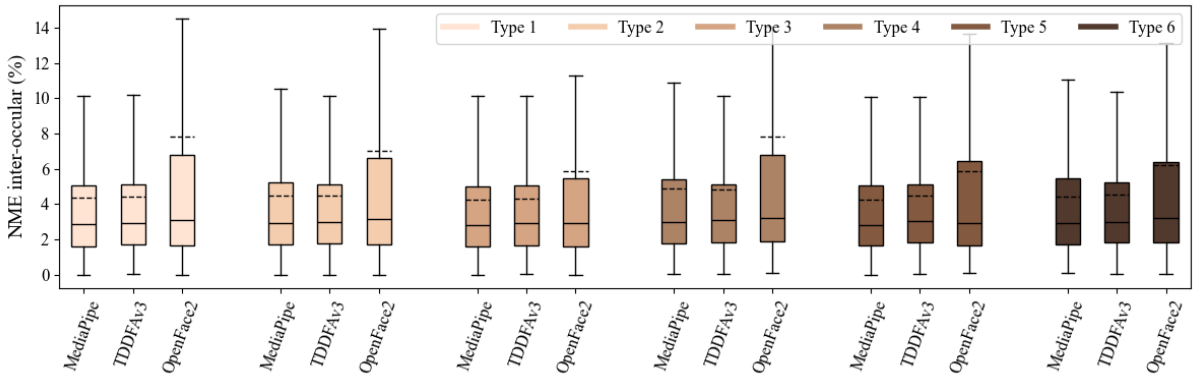


Fig. 2. Algorithm NME by Fitzpatrick skin types.

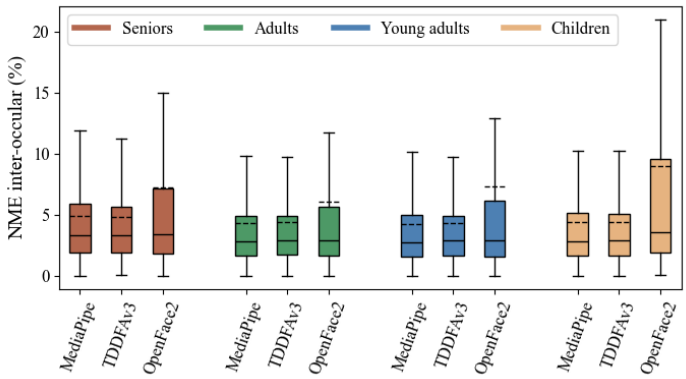


Fig. 3. Algorithm NME by age groups.

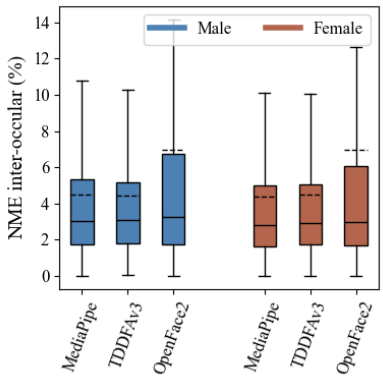


Fig. 4. Algorithm NME by sex.

TABLE I. NUMBER OF KEYPOINT AFFECTED BY FACTOR VARIATIONS

Algorithm	Age	Skin tone	Sex	Occlusion	Lighting	Expression
MediaPipe	7	2	6	2	2	3
3DDFAv3	8	4	7	1	3	4
OpenFace2	10	0	1	0	5	6

To determine the effect of different demographics on statistically significant keypoints, post hoc analyses were performed. The general trends from this analysis are as follows: For the affected keypoints, the largest increase in NME due to a specific demographic relative to the mean per algorithm was 35% for MediaPipe, 36% for 3DDFAv3, and 48% for OpenFace2. Regarding age, all algorithms performed well on adults and young adults. However, seniors led to higher NME for both MediaPipe and 3DDFAv3, while children caused the biggest increase for OpenFace2. The impact of sex was less clear; MediaPipe performed slightly better on women (NME decrease of 0.003), while 3DDFAv3 showed keypoint variations in both directions. For skin tone, Mediapipe and 3DDFAv3 indicated a slight advantage for lighter types, with Mediapipe performing better on types 1 and 2 than 4 and 3DDFA performing worse on types 4, 5 and 6 than on types 1, 2 and 3.

IV. DISCUSSION & CONCLUSION

The first objective was to test for demographic biases within three popular algorithms. Analysis revealed a significant impact on specific keypoints, with errors relative to the mean increasing by up to 48%. Even though only a small subset of keypoints showed significance, their impact was comparable to other known problematic factors, such as occlusion or facial expression. However, it is important to note that FAIRSET was not designed to be balanced for these additional factors. Age affected the highest number of keypoints compared to other factors. Therefore, when working with age groups such as children or seniors, users should expect worse results compared to adults. To our knowledge, no other studies have examined the direct link between age and the performance of face landmarking algorithms. However, an emotion analysis study showed similar results, where the estimations performed on the expressions of older people performed worse compared to young adults [14]. Several keypoints were affected by sex, though this effect varied from algorithm to algorithm. OpenFace2 only had one affected keypoint, while 3DDFAv3 had many, swaying in favor of both men and women. As stated above, MediaPipe performed slightly better on women, with an NME reduction of 0.003. Although small, this difference matches the results

provided for their dataset [15]. The takeaway is that sex does affect accuracy, but in an unpredictable manner, depending on the algorithm and the specific keypoint. Skin tone affected the smallest number of keypoints for each algorithm. OpenFace2, which generally has more variable results, did not show statistical significance in this factor. A general trend was observed that darker skin tones performed worse, but only one or two keypoints per skin tone displayed this. This does not match the results provided by MediaPipe, since type 4, which performed the worst on FAIRSET, was MediaPipe's second best skin type [15]. This discrepancy could be due to the relatively small sample size for darker skin tones in FAIRSET. All ANOVA assumptions except normality were met, so this difference could be attributed to the dispersion of images influenced by environmental factors within skin tone subgroups. This same issue may also have occurred during the age and sex analyses. Future work will focus on analyzing this dispersion in greater detail to understand its impact on error rates between age, skin tone, and sex subgroups. This study found a statistically significant link with skin tones for two of the three algorithms, but further analysis is required to make safe assertions.

The second objective aims to propose guidelines for selecting an appropriate face landmarking algorithm. Evidently, the choice depends on the specific use case. Our results suggest that when studying specific demographics, users should validate how their envisaged algorithm performs on this group, as results may vary. Among the observed algorithms, users should avoid OpenFace2, as it is generally less accurate despite displaying fewer statistical links with demographics. When studying problematic demographics such as seniors, there is no clear winner between MediaPipe and 3DDFAv3, as both showed that specific keypoints performed worse on seniors than other age groups, with similar NMEs. On the other hand, the effect of sex varied a lot between the two algorithms, with MediaPipe performing better on women, while 3DDFAv3 swayed both ways depending on the keypoint. This highlights the importance of validating an algorithm's performance across demographics through statistical analysis, as demonstrated in this paper. The effect a specific demographic on an algorithm remains undocumented and unpredictable. Therefore, using a dataset like FAIRSET, whether during development or before selecting an algorithm, will guide users toward a choice that better fits their needs.

Acknowledgment

This manuscript was reviewed for language and fluidity with the assistance of Microsoft Copilot. This study is supported by the Canada Research Chair in Biomechanical Movement Signature to Better Understand and Reduce Disability.

References

[1] F. Danieau, I. Gubins, N. Olivier, O. Dumas, B. Denis, T. Lopez, N. Mollet, B. Frager, and Q. Avril, "Automatic Generation and Stylization of 3D Facial Rigs," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Mar. 2019, pp. 784–792, iSSN: 2642-5254. [Online]. Available: https://ieeexplore.ieee.org/document/8798208

[2] B. Al-Baker, A. Alkalaly, A. Ayoub, X. Ju, and P. Mossey, "Accuracy and reliability of automated three-dimensional facial landmarking in medical and biological studies. A systematic review," *European Journal of Orthodontics*, vol. 45, no. 4, pp. 382–395, Jul. 2023.

[3] Y. Wu and Q. Ji, "Facial Landmark Detection: A Literature Survey," *International Journal of Computer Vision*, vol. 127, no. 2, pp. 115–142, Feb. 2019. [Online]. Available: https://doi.org/10.1007/s11263-018-1097-z

[4] A. Khalil, S. G. Ahmed, A. M. Khattak, and N. Al-Qirim, "Investigating Bias in Facial Analysis Systems: A Systematic Review," *IEEE Access*, vol. 8, pp. 130 751–130 761, 2020, conference Name: IEEE Access. [Online]. Available: https://ieeexplore.ieee.org/document/9130131

[5] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 Faces In-The-Wild Challenge: database and results," *Image and Vision Computing*, vol. 47, pp. 3–18, Mar. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0262885616000147

[6] S. Escalera, M. T. Torres, B. Martínez, X. Baró, H. J. Escalante, I. Guyon, G. Tzimiropoulos, C. Corneanu, M. Oliu, M. A. Bagheri, and M. Valstar, "ChaLearn Looking at People and Faces of the World: Face AnalysisWorkshop and Challenge 2016," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2016, pp. 706–713, iSSN: 2160-7516. [Online]. Available: https://ieeexplore.ieee.org/document/7789583

[7] Z. Zhang, Y. Song, and H. Qi, "Age Progression/Regression by Conditional Adversarial Autoencoder," Mar. 2017, arXiv:1702.08423 [cs]. [Online]. Available: http://arxiv.org/abs/1702.08423

[8] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 Faces in-the-Wild Challenge: The First Facial Landmark Localization Challenge," in *2013 IEEE International Conference on Computer Vision Workshops*. Sydney, Australia: IEEE, Dec. 2013, pp. 397–403. [Online]. Available: http://ieeexplore.ieee.org/document/6755925/

[9] N. Zelovic, I.-M. Joly, E. Riesco, and K. Lebel, "Towards a Gold Standard for AI Facial Landmarks Estimation: Constructing a Balanced and Inclusive Landmark Database," in *Proceedings of the 22nd IGS conference: Investigating Human Movements: Handwriting and Beyond*. Montréal, Canada: Presses Internationales Polytechnique, 2025.

[10] I. Grishchenko, A. Ablavatski, Y. Kartynnik, K. Raveendran, and M. Grundmann, "Attention Mesh: High-fidelity Face Mesh Prediction in Real-time," Jun. 2020, arXiv:2006.10962 [cs]. [Online]. Available: http://arxiv.org/abs/2006.10962

[11] Z. Wang, X. Zhu, T. Zhang, B. Wang, and Z. Lei, "3D Face Reconstruction with the Geometric Guidance of Facial Part Segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 1672–1682.

[12] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "OpenFace 2.0: Facial Behavior Analysis Toolkit," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. Xi'an, China: IEEE Press, May 2018, pp. 59–66. [Online]. Available: https://doi.org/10.1109/FG.2018.00019

[13] K., Kanggeon, T. Baltrušaitis, A. Zadeh, L.-P. Morency, and G. Medioni, "Holistically Constrained Local Model: Going Beyond Frontal Poses for Facial Landmark Detection," in *Proceedings of the British Machine Vision Conference (BMVC)*, R. C. Wilson, E. R. Hancock, E. R., and W. A. P. Smith, Eds. BMVA Press, Sep. 2016, pp. 95.1-95.12. [Online]. Available: https://dx.doi.org/10.5244/C.30.95

[14] E. Kim, D. Bryant, D. Srikanth, and A. Howard, "Age Bias in Emotion Detection: An Analysis of Facial Emotion Recognition Performance on Young, Middle-Aged, and Older Adults," in Proceedings of the 2021 *AAAI/ACM Conference on AI, Ethics, and Society*, ser. AIES '21. New York, NY, USA: Association for Computing Machinery, Jul. 2021, pp. 638–644. [Online]. Available: https://doi.org/10.1145/3461702.3462609

[15] G. AI, "Face landmark detection guide| Google AI Edge" Nov. 2024 [Online] Available: https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker

Improving Personality Prediction in Handwriting Analysis through Optimal Input Granularity

Yan Xu*
CU & SDNU
Montreal, Canada & Jinan, China
y_xu2@encs.concordia.ca

Yufang Tang*
Shandong Normal University(SDNU)
Jinan, Shandong 250014, China
tangyufang@outlook.com

Ching Y. Suen
Concordia University(CU)
Montreal, Quebec H3G 1M8, Canada

***Abstract*— This study explores the optimization of input sample granularity to improve deep learning model performance in handwriting analysis for personality prediction. Traditionally, adapting deep learning models to specific application scenarios focuses on fine-tuning internal parameters, often overlooking potential optimizations in input samples. We propose a novel approach where the characteristics of input samples, particularly split patterns and resolution choices, can significantly impact model classification performance. To this end, we introduce the concept of optimal granularity, determined by two key factors—split pattern and resolution (resize rate)—which influence the structure of both training and testing datasets.

Through experiments on a custom-built handwriting dataset, we trained two deep learning models to predict personality traits such as extraversion (EXT) and conscientiousness (CON). Our results show that the optimal granularity for both traits is achieved with a 3x8 split pattern and a 0.3 resize rate, where each sample contains 3 to 4 lines of text with approximately 4 to 6 words per line, yielding the best performance across various handwriting styles.

This work not only provides actionable strategies for handwriting-based personality analysis but also lays the groundwork for optimizing input samples in deep learning applications across other domains. Future research will explore additional factors influencing granularity and aim to improve model accuracy and robustness through fine-tuning parameters.

Keywords*— *Big Five, personality prediction, optimal granularity, split pattern, resolution, deep learning

I. Introduction

Handwriting analysis, also known as graphology, is used to uncover relationships between graphological features and various personality traits, such as character, gender, age, occupation, experience, and disease. This technique has applications in healthcare, recruitment, criminal identification, education, and more [1].

As a classification problem, handwriting analysis can employ either classical machine learning methods or deep learning methods. Deep learning methods, with their more complex structures, can extract more informative features to analyze personality traits. However, deep learning models are often regarded as black-box systems that offer limited interpretability regarding their training dynamics and decision-making processes. As such, a rigorous evaluation of input data is imperative to ensure reliable and meaningful outcomes.

Most data sets used in personality prediction research are sourced from public or private collections, typically focusing on a single language, such as English. The size of private data sets can vary significantly, ranging from as few as fifty samples [2] to nearly two thousand samples [3].

Despite this, most research does not clearly specify the exact size of the input samples after preprocessing I [4]. Some samples may be whole, while others may be divided, but the exact size of the division is often not mentioned. In this paper, we delve deeper into this issue to provide clearer guidelines for future research.

II. Method

We hypothesize that varying input sizes can lead to different outcomes in model performance [1]. Here, we define input sizes as the different spatial crops or regions extracted from the original input image, which are used to provide the model with varying contextual or localized information. For example, when the entire sample is used as input, the model is likely to capture more global information. In contrast, when only a portion of the sample is used, the model tends to focus on more localized information.

A. Dataset

We used a proprietary data set collected in 2019, which contains samples in multiple languages [5]. This data set includes 234 samples, each annotated with labels based on psychological questionnaires (objective measurements) and graphologist evaluations (subjective measurements). For the model prediction standard, we focused on the Big Five personality traits [6], which are recognized for their long-term stability.

Our labeling method combines inputs from both psychological questionnaires and graphologist evaluations. The labels are: 1 ”Very Low (VL)”, 2 ”Low”, 3 ”Average (Avg)”, 4 ”High”, and 5 ”Very High (VH)”. If only one source provides a label, that value is used. When one source gives an ”Avg” rating and the other a ”High”, we assign the higher value (”High”). In cases where one rating is ”Low” and the other is ”High”, we select the average value (”Avg”). This approach is analogous to considering the evaluations of two independent reviewers and arriving at a balanced score, for instance, by computing the

average. Each category we defined is associated with a corresponding numerical value.For discrepancies between ”VL” and ”High”, we compute the actual midpoint to decide if the final label should be ”Low” or ”Avg”.

Based on the distribution of each label for two classes: Extraversion (EXT) and Conscientiousness (CON), we found that the label ”VL” has only one sample in the EXT class and none in the CON class, so we opted to exclude it. As a result, each class has four remaining labels: 2 (Low), 3 (Avg), 4 (High), and 5 (VH), without representation of the VL category. The models are inherently incapable of predicting VL outcomes, which constitutes a limitation of this study.

B. Granularity

When using ”part of the sample” as input, it’s important to determine the optimal segment size. Granularity refers to the level or degree of segmentation and involves two parameters: split pattern and resolution.

1) Split Pattern: Global and local information play crucial roles in multi-scale learning, a machine learning approach where a model processes input data at various levels of granularity to capture both overarching patterns and finer details.

In handwriting analysis, these two types of information are equally essential. Global information provides insights into large-scale patterns and overarching context, such as the overall structure and style of handwriting. In contrast, local information focuses on fine-grained details within specific regions, such as letter movements, stroke patterns, textures, and edges.

A 1x1 split means no segmentation, using the entire sample as-is, while a 1x2 split divides the sample into 1 column and 2 rows, effectively doubling the sample count.

To examine a wide range of splitting patterns, we tested various combinations of columns and rows, including: 1x2, 1x4, 1x6, 1x8, 1x10, 3x4, 3x8, 6x8, 6x16, 9x12, 9x24, 2x1, 4x1, 6x1, 8x1, 10x1, 4x3, 8x3, 8x6, 16x6, 12x9, and 24x9. The different split patterns yield a variety of segment shapes, including square-like sections, vertical strip matrices, and horizontal strip matrices.

As samples are divided, they may exhibit different distributions of text and spacing, resulting in a range of segment sizes and quantities, see Table I [4].

TABLE I: Number of each label (Avg, High, Low, VH) for two classes (EXT/CON).

Split by col. and rows	Num. of Avg (EXT/CON)	Num. of High (EXT/CON)	Num. of Low (EXT/CON)	Num. of VH (EXT/CON)	EXT (Train/Test)	CON (Train/Test)
1x2	104/70	36/78	17/6	2/5	13/3	4/1
1x4	228/167	81/159	40/14	4/13	32/6	10/2
1x6	377/288	125/249	70/21	5/19	56/11	15/3
1x8	513/376	162/343	91/28	7/27	72/15	21/4
1x10	645/488	205/428	125/36	9/33	100/20	26/5
2x1	98/70	34/72	20/6	2/6	16/3	4/1
3x4	718/526	225/481	137/45	11/41	109/22	32/7
3x8	1538/1125	460/1036	306/88	22/81	244/49	64/13
4x1	215/153	76/151	38/15	4/14	30/6	11/2
4x3	679/486	219/475	126/43	11/33	100/20	26/5
6x1	343/249	118/231	57/22	6/22	45/9	17/4
6x8	3183/2305	922/2178	668/181	43/169	534/107	135/27
6x16	6343/4611	1824/4339	1343/359	85/315	1074/215	252/50
8x1	477/343	161/322	80/32	8/29	64/12	23/4
8x3	1478/1058	451/1025	284/87	23/71	227/45	56/12
8x6	3161/2354	942/2153	676/176	42/155	540/108	124/24
9x12	7230/5289	2075/4949	1569/402	91/366	1255/251	292/59
9x24	14618/10598	4069/10037	3269/822	185/769	2615/523	615/123
10x1	600/432	202/416	109/36	10/37	87/17	28/6
12x9	7318/5315	2079/5007	1569/408	97/370	1255/251	296/59
16x6	6411/4755	1880/4392	1419/356	83/321	1135/227	256/52
24x9	14325/10417	4031/9833	3142/772	181/731	2513/503	584/117

In Table I, the first column splits the input sample by columns and rows. The second through fifth columns show the count of each label (Avg, High, Low, and VH), with the number for class EXT appearing before the ”/” symbol and the number for class CON after it. The sixth and seventh columns display the counts for classes EXT and CON, with the number of training samples shown before the ”/” symbol and the number of test samples after it.

As the split grows, the sample count increases from tens to tens of thousands. However, each label within a class shows imbalance. To mitigate this, we use a strategy of selecting the minimum number for each label within a class.

In the EXT class, the label ”VH” is significantly underrepresented compared to the others, so we exclude it from the EXT analysis. Thus, the EXT class includes three labels: Avg, High, and Low. In contrast, the CON class retains four labels: Avg, High, Low, and VH.

2) Resolution: Once we've determined the optimal split for the samples, we need to decide the final size of the input sample for the training model—much like adjusting a magnifying glass to focus on the segmented sample.

For each split, we ideally aim to apply a range of resize rates, from 1.0, 0.9, and 0.8 down to 0.1. However, not every split can support all these resize rates. This limitation arises either from computational constraints when the number of loaded training samples becomes too large, or from the model's requirements, as the input sample size for the training model ConvNeXtTiny [7] [8] [9] must be at least (32, 32). For instance, the 1x2 split is feasible only at a resize rate of 0.2; higher rates like 0.3 or more exceed our machine's processing capacity(an Intel Core i7-11700 CPU @ 2.50GHz, 32 GB RAM, and an NVIDIA GeForce RTX 3080 GPU (10 GB VRAM)??). Even with the same sample size (1020, 660), the CON class can be trained with both loss and accuracy metrics, but the EXT class cannot, due to differing label counts between the two classes.

We independently trained two models for the EXT and CON classes of the Big Five personality traits, applying two granularity parameters: split pattern and resolution (resize rate).

In the calculation, there may have some unavailable values, which is due to either the computational limit when the number of loaded training samples is too large, or the model's limitations, as the input sample size for ConvNeXtTiny must be at least (32, 32).

III. Experiment

A. Settings

Our experimental setup was organized as follows: the dataset was split into training, validation, and testing sets in a 6:2:2 ratio. We used a batch size of 16 and trained the model over 100 epochs. Categorical cross-entropy was applied as the loss function, with the Adam optimizer for optimization, the accuracy is balanced accuracy. For classification, we employed the ConvNeXtTiny model, which outperformed the other seven commonly used classification models [7] in the context of handwriting-based personality trait analysis, achieving the highest overall performance. Although regression models are capable of reflecting the extent of prediction deviations, the focus of this study is on the accurate identification of categorical rating levels. Accordingly, we employ a classification-based modeling strategy.

B. Analysis

Overall, the EXT class contains three to four times more samples than the CON class, which would suggest that higher accuracy is expected for the EXT class. Indeed, the EXT class achieved the highest accuracy of 0.7778, while the CON class reached a good accuracy of 0.75. Although the difference between their best accuracies is only 0.0278, this result further indicates that using fewer training samples can still yield competitive performance.

For both classes, the best splits—with more than half of the accuracy values above 0.6—are 3x8 (with widths ranging from 510 to 850 and heights from 247 to 412) and 8x6 (with widths from 191 to 445 and heights from 330 to 770). As shown in Fig. 1, the 3x8 split includes 3 to 4 lines of text, with each line containing 4 to 6 words, resulting in a total word count between 12 and 24. The 8x6 split contains 3 to 5 lines of text, with each line containing 2 to 3 words, totaling between 6 and 15 words. Therefore, when selecting a sample, it is ideal to have at least 3 to 4 lines, with approximately 4 words per line, for a total word count between 12 and 16.

Splitting the data is essential, but more splits do not always lead to better results—the ideal split must be balanced. With the proper granularity, a large number of training and test samples is not necessary.

IV. Conclusion

This paper presents empirical evidence supporting our proposed approach, showing that the optimal granularity for handwriting analysis is not simply a matter of increasing the number of splits. Rather, it depends on selecting a split pattern and resolution that jointly capture both global and local features of the input.

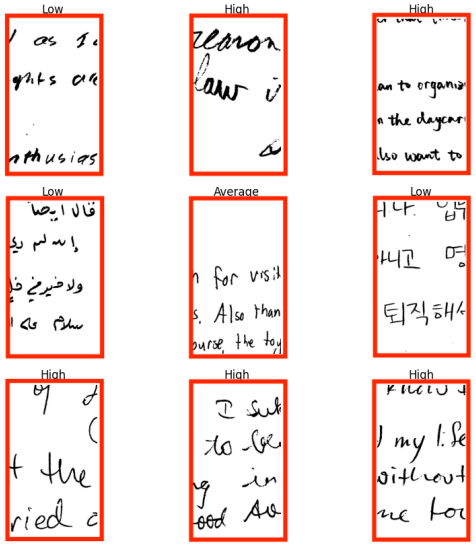
Through extensive experimentation, we found that a 3×8 horizontal split combined with a resize rate of 0.3—corresponding to segments containing approximately 3 to 4 lines of text and around 4 words per line—achieved the best overall performance across various handwriting styles. This configuration effectively balances structural detail with contextual coherence, highlighting the importance of jointly tuning spatial partitioning and input scaling.

These findings underscore the need for thoughtful design of input granularity in handwriting-based personality analysis and suggest that arbitrary increases in resolution or segmentation may not yield improved outcomes.

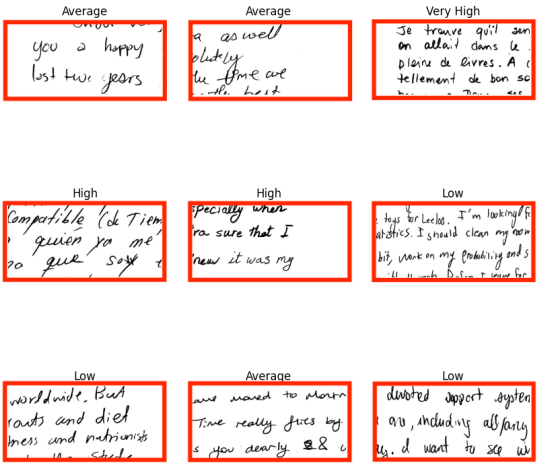
For future work, we plan to investigate additional factors influencing granularity, such as language-specific characteristics or stroke-level segmentation strategies. We also aim to further improve model accuracy and reduce training loss through fine-grained parameter tuning and focused analysis within a single language domain.



(a) EXT 3x8



(b) EXT 8x6



(c) CON 3x8



(d) CON 8x6

Fig. 1: Split samples for both classes.

ACKNOWLEDGMENT

This study was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] Y. Xu, Y. Tang, C. Y. Suen, "Split or Not to Predict Big Five Personality?," The 6th Asia Conference on Cognitive Engineering and Intelligent Interaction (CEII 2023). Hong Kong, China, 15-16 December 2023.

[2] A. Chitlangia and G. Malathi, "Handwriting analysis based on histogram of oriented gradient for predicting personality traits using svm," Procedia computer science, vol. 165, pp. 384–390, 2019.

[3] P. Joshi, P. Ghaskadbi, and S. Tendulkar, “A machine learning approach to employ- ability evaluation using handwriting analysis,” in Proceedings of the second International Conference on Advanced Informatics for Computing Research (ICAICR), Shimla, India, July 14–15, 2018, Revised Selected Papers, Part I 2, Springer, 2019, pp. 253–263.
[4] Y. Xu, Y. Tang, C. Y. Suen, “Determining Optimal Granularity for Effective Handwriting Analysis”. The Workshop Image Mining, Theory and Applications(IMTA- IX) of the 27th International Conference on Pattern Recognition(ICPR). Kolkata, India, 1-5 December 2024. Published on Special Issue “Proceedings of IMTA-IX-202”, Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications. 2024: 34(4).
[5] Y. Xu, Y. Tang, C. Y. Suen, “Review of Handwriting Analysis for Predicting Personality Traits,” IAPR Joint International Workshops on Statistical Techniques in Pattern Recognition and Structural and Syntactic Pattern Recognition (S+SSPR 2022). Montreal, Canada, 26-27 August 2022, LNCS: 54–63.
[6] W. T. Norman, “Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings,” The journal of abnormal and social psychology, vol. 66, no. 6, pp. 574–583, 1963.
[7] Y. Xu, Y. Tang, C. Y. Suen, “Two Key Factors in Handwriting Analysis for Personality Prediction,” The 5th International Conference on Image, Video Processing and Artificial Intelligence (IVPAI 2023). Shenzhen, China, 14-15 December 2023, SPIE, 2024, 13074: 102–107.
[8] Y. Xu, Y. Tang, C. Y. Suen, “Unlocking Raw Data: The Power of Micro-Inspection in Handwriting Analysis,” The 4th International Conference on Pattern Recognition and Artificial Intelligence (ICPRAI). Jeju Island, Korea, 3-6 July 2024, in press.
[9] Liu, Z., Mao, H., Wu, C. Y., et al.,“A convnet for the 2020s”. In Proceedings of the IEEECVF Conference on Computer Vision and Pattern Recognition, pp. 11976-11986 (2022).

Comparing Character Shape between Two Japanese Handwriting Databases of Forty-Years' Difference

Yoshinori Akao
National Research Institute of Police Science, 6-3-1, Kashiwanoha, Kashiwa, Chiba 2770882, JAPAN
akao@nrips.go.jp

***Abstract*—** Averaged shape of handwriting was defined as kernel density estimation of strokes and evaluated for selecting reference character shape in forensic handwriting examination. The difference of averaged character shape calculated by two different Japanese handwriting databases, which was collected in different era of forty-years-difference. Experimental results showed several changes between two databases. Increasing or decreasing in writing strokes, connection status between strokes, direction or length of strokes, proportion of parts in character, slant and overall shapes are examples of these differences. The importance of using proper handwriting database suited to samples for analysis was suggested.

Keywords—character shape, Japanese handwriting, averaged shape, kernel density estimation

I. Introduction

Averaged shape of handwriting gives us crucial for selecting reference character shape in forensic handwriting examination. Kernel density estimation of handwriting stroke is one of the approaches, and it became like the font shape appeared in the textbook used in school [1]. However, the effects due to the kind of databased used for calculation were not considered. Therefore in this study, the author investigated the difference of averaged character shape calculated by two different Japanese handwriting databases, which was collected in different era of forty-years-difference.

II. Materials and Method

Sample for analysis were two handwriting databases. One was collected in 1978 [2], and another was in 2019 [3]. The kinds of character used in this experiment were 370 (Chinese "Kanji" 314, Japanese "Hiragana" 46 and 10 numerals). Each character was written in bounding box with the size of 10 mm by 10 mm using ball point pen. Number of participants were 60. Each participants wrote these characters five times. The kinds of characters, number of participants and number of repetition were different between two databases. But these common 370 kinds of characters were selected. The number of participants (60) and the number of repetition (5) were also made to be common.

Each character was digitized into grayscale image with 300 by 300 pixels. Kernel density of handwriting stroke of each kind of character was calculated for each writer. The size of character was normalized to be the second order moment of the image equal to 50 before the calculation of kernel density. After that, kernel density images of each writer were averaged. The author defined that kernel density image as "averaged shape" of each type of character.

The averaged shape was calculated for both database 1978 and 2019. The shape was evaluated with the observation of human eye by the author of this study.

III. Results and Discussion

A. Chinese "Kanji" characters

Figure 1 shows an example of comparing averaged character shape of Chinese "Kanji" characters between 1978 database and 2019 database. The fundamental structure were common and read properly considering the purpose of characters. But fine differences were found in various aspects in characters, such as overall shapes, ration between parts or strokes, slant of each stroke. The way of pen movement at the end of stroke was also different in fine scale.

B. Japanese "Hiragana" characters

Figure 2 shows an example of comparing averaged character shape of Japanese "Hiragana" characters between 1978 database and 2019 database. Same as the case of Chinese "Kanji" characters, the fundamental structure were common and read properly. Japanese "Hiragana" character is more cursive than Chinese "Kanji". So it makes more effect connected to the way of pen movement. The difference between two databases varied according to the kind of characters. Slant, proportion between parts and degree of curvature in strokes were examples of the difference.

C. Numerals

Figure 3 shows an example of comparing averaged character shape of numerals between 1978 database and 2019 database. The fundamental structure were common and read properly in the case of numerals also. The biggest difference was slant. In 1978 database, characters were more slanted than in 2019 database. The proportion between parts in each character was also different.

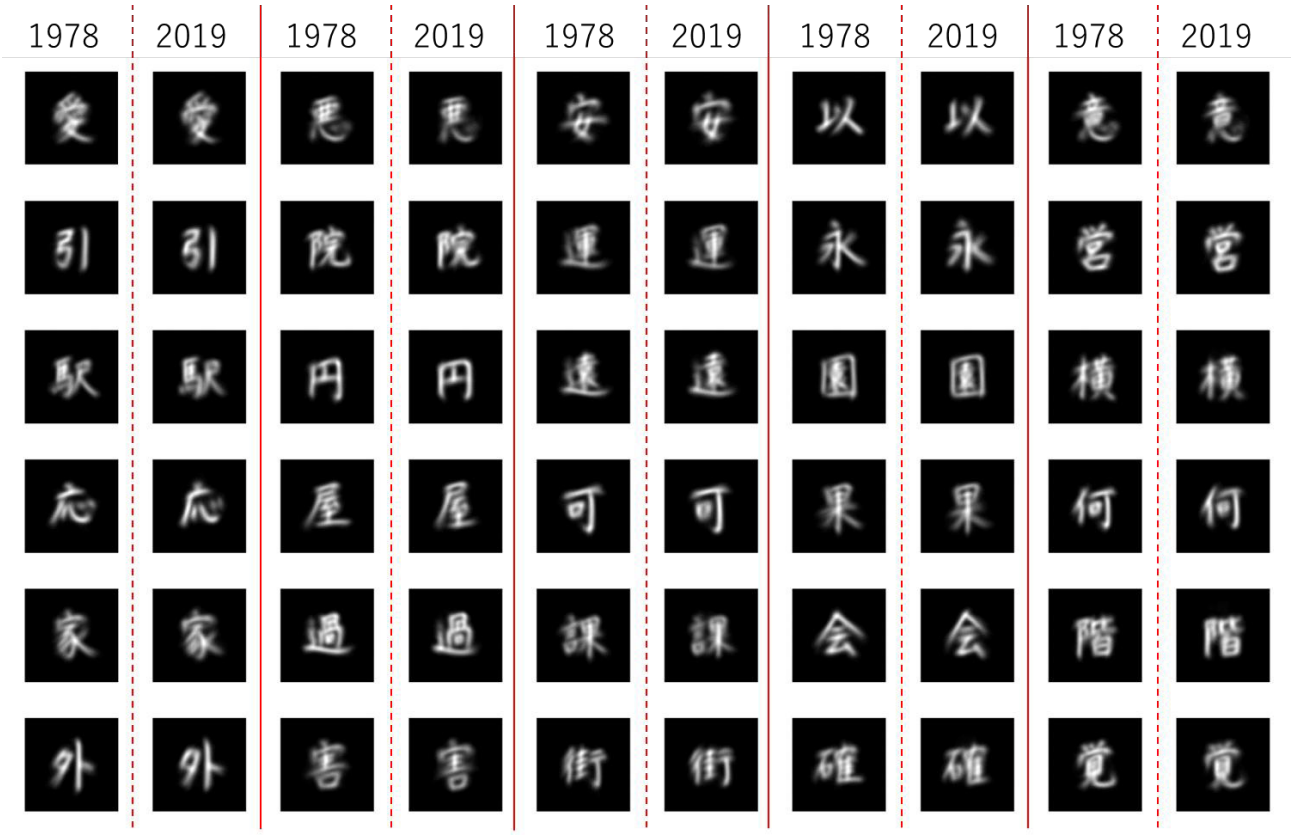


Fig. 1. Comparison of Chinese “Kanji” characteres (first 30 characters).

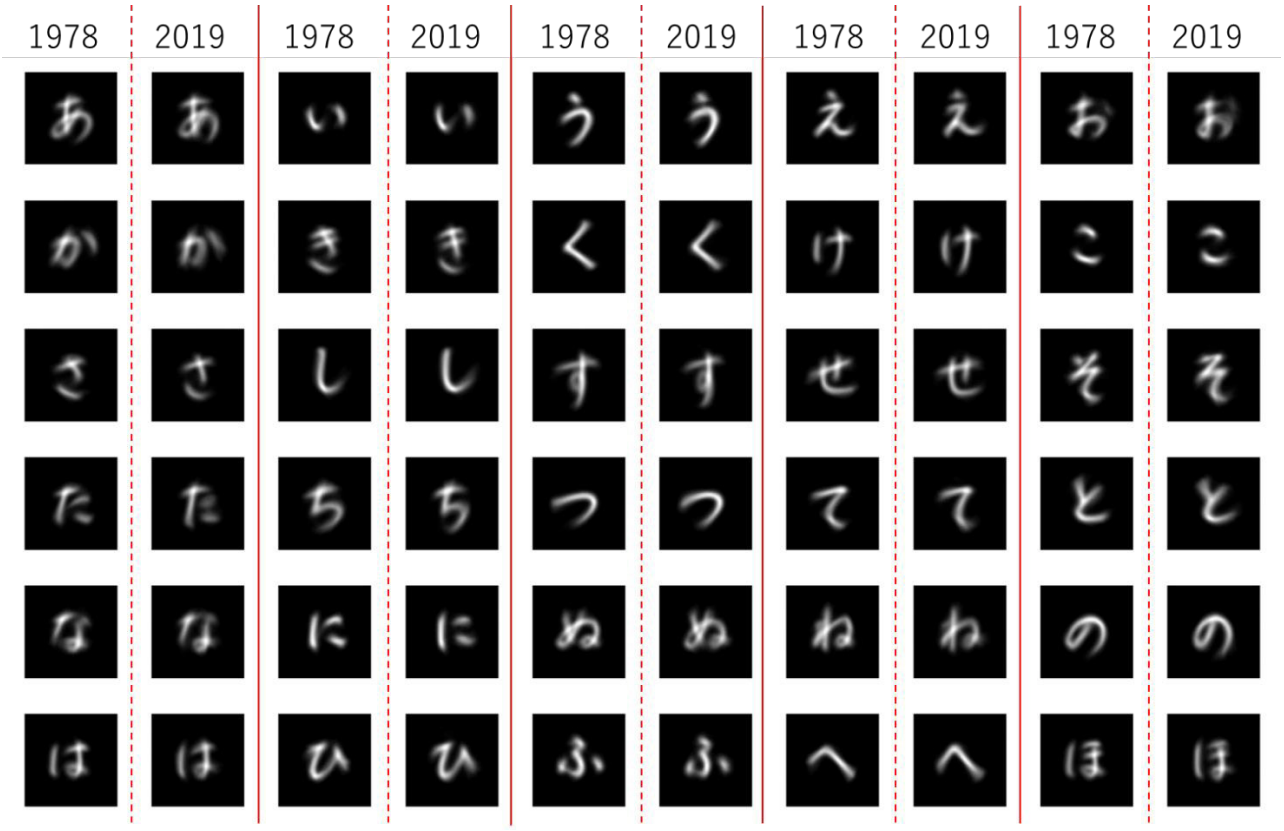


Fig. 2. Comparison of Japanese “Hiragana” characters (first 30 characters).

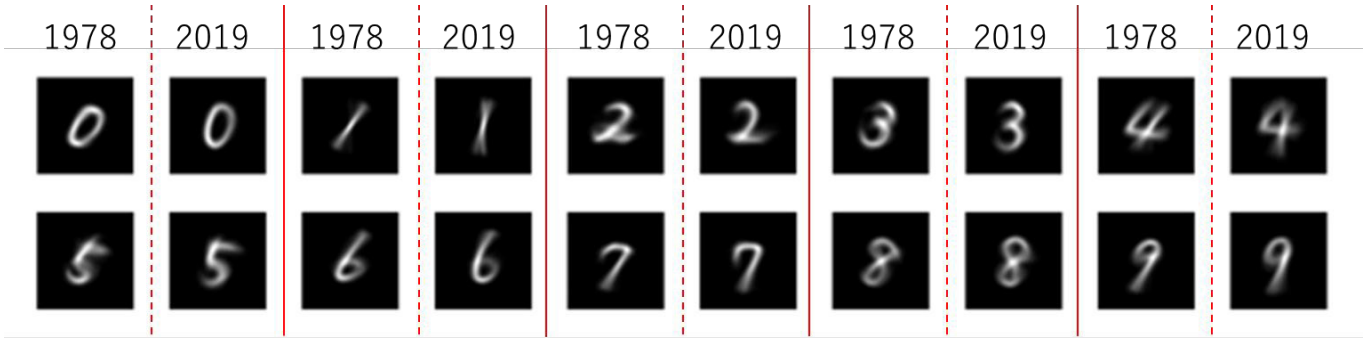


Fig. 3. Comparison of numerals.

D. Characteristic examples

Figure 4 shows characteristic examples of the difference between two databases.

In 1978 database, strokes of Japanese Hiragana "そ" were composed of two parts but it was one in 2019 (Fig.4(a)). It showed the character was formerly tended to be written in two strokes, but it changed to be written in one stroke recently.

Numeral "4" drastically changed its shape between two databases. The top part was separated in 1978, but it became connected in 2019 (Fig.4(b)). It was considered that the shape was more fitted to that of textbooks used in Japanese elementally schools in 1978, but it gradually affected by the shape of fonts used in printed materials and computers.

For Chinese "Kanji" characters, the following tendency were observed. Firstly, strokes that goes from top to bottom-leftward tended to be long and more leftward in 1978, but it became shorter and more vertical in 2019. Secondly, the end part of strokes that goes from top to bottom were jumped toward left-upward in 1978, but it became stopped or the amount of the jump decreased. Thirdly, the shape of components "口", "日" and "田" appeared in the characters were different. In the database of 1978, their vertical strokes were shorter than that of horizontal strokes. But the difference in length between the vertical and horizontal strokes became smaller or equal in 2019. Consequently, its overall shape became more squarely (Fig.4(c)). The background of these changes were considered to be stemmed from calligraphy using brush in education. The opportunity or experiences using brush in writing characters were drastically decreased in this period. The characteristic features listed above were reflecting that of calligraphy using brush.

Another interesting example was Chinese character "必" (Fig.4(d)). The character shape was similar to another Chinese character "心" in 1978, but it has drastically changed into square shaped in 2019. The author considered the background of this issue as follows. It was formerly taught in elementary school to write with stroke order similar to that of another Chinese character "心". But the stroke order in the education was changed in this period.

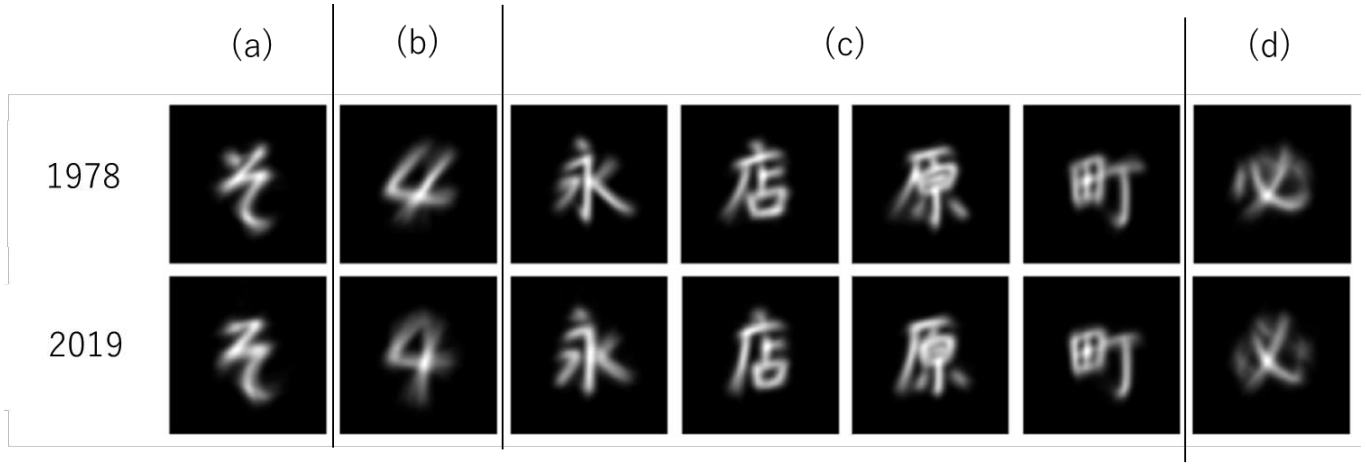


Fig. 4. Characteristic examples, (a) Japanese Hiragana "そ", (b) numeral "4", (c) Chinese characters and (d) Chinese character "心".

IV. Summary

Averaged character shapes were different between two databases forty-years-difference. The difference was considered to reflect the change of writing manner in forty years. The experimental results suggested the importance of referring proper handwriting database. That is to say, the age and population should be suited to the sample for database analysis in forensic handwriting examination. Other factors such as instructions for participants, the effect of writing devices, ages or educations the participants received might affect the results. Further investigation should be followed in future works.

Acknowledgment

Part of this work was supported by JSPS KAKENHI Grant Number 21K12577.

References

[1] Y. Akao *et al.*, Japanese Journal of Forensic Science and. Technology, 20(Suppl.), 151 (2014).

[2] Document examination sedtion of National Research Instituge of Police Science, Reports of National Research Institute of Police Science, 31, 212 (1978).

[3] Y. Seki *et al.*, Reports of National Research Institute of Police Science, 69, 33 (2020).

Development of an Automated Tool for the Extraction of Muscle Synergies from Electromyographic Data

Flore LE BLANC
Université de Sherbrooke, Centre de recherche sur le vieillissement
1036 rue Belvédère Sud, Sherbrooke QC J1H 4C4
flore.le.blanc@usherbrooke.ca

Simon BRIÈRE
Université de Sherbrooke, Centre de recherche sur le vieillissement
1036 rue Belvédère Sud, Sherbrooke QC J1H 4C4
simon.briere@usherbrooke.ca

Daniel BOURBONNAIS
Centre de recherche interdisciplinaire en réadaptation du Montréal-Métropolitain
Institut de réadaptation Gingras-Lindsay-de-Montréal
6300, Avenue Darlington, Montréal (Québec) H3S 2J4
daniel.bourbonnais@umontreal.ca

Camille MARTIN
Université de Sherbrooke, Centre de recherche sur le vieillissement
1036 rue Belvédère Sud, Sherbrooke QC J1H 4C4
camille.martin3@usherbrooke.ca

Sylvie LAFRENAYE
Centre de Recherche du CHUS – Département de Pédiatrie
3001 12 Avenue Nord, Sherbrooke, Quebec J1H 5H3
sylvie.lafrenaye@usherbrooke.ca

Guillaume LÉONARD
Université de Sherbrooke, Centre de recherche sur le vieillissement
1036 rue Belvédère Sud, Sherbrooke QC J1H 4C4
guillaume.leonard2@usherbrooke.ca

Abstract

The aim of this study is to test a tool developed to automate the pre-processing and analysis of muscle coordination during a locomotion task as well as identify potential pain-induced differences in muscle coordination during gait with respect to pain-related psychological beliefs.
Developed by Dr. Daniel Bourbonnais and colleagues, this tool utilizes the non-negative matrix factorization method to extract muscle co-activation patterns and compare these activations between trials and participants. We used an experimental pain paradigm in which participants were equipped with EMG sensors and instructed to walk on a split-belt treadmill in a pain-free condition and a painful condition. Acute pain was induced on the dominant leg using 1% capsaicin cream. We also measured psychological factors relating to pain using the Tampa Scale of Kinesiophobia and the Pain Catastrophizing Scale to explore correlations between these scores and behavioural patterns in pain avoidance.
Our project proposes a non-invasive, streamlined way to analyze synergistic patterns of muscle activation during locomotion tasks. This approach will enable us to investigate how these patterns are influenced by pain, kinesiophobia and pain catastrophizing. We anticipate that our results will reveal an interaction between kinesiophobia, pain catastrophizing, and the degree to which muscle synergies are affected by the presence of acute experimental pain.
Our aim is twofold: firstly, to evaluate the reliability of this new approach for evaluating muscle synergies; and secondly, to determine whether this new approach can be used to investigate pain-induced changes in motor control and synergies. We hope to make this tool accessible for public use, making this type of analysis more useful for researchers and clinicians alike in the investigation of musculoskeletal and neurodegenerative pathologies.

Keywords—movement, locomotion, muscle synergies, clinical transferability, rehabilitation, electromyography

I. Introduction

Understanding muscle coordination during movement is critical for applications in rehabilitation, assistive device development, and human performance optimization. One approach to studying motor control is through the analysis of muscle synergies (MS) [1]. MS represent the coordinated activation of groups of muscles contributing to a particular movement. A single muscle may participate in multiple synergies, and each synergy can activate several muscles. MS can be identified by factor analysis, offering insights into neuromuscular control strategies. A widely used technique in this process is non-negative matrix factorization (NNMF) (Gamma model based on J divergence) [2]. Although several methods can be used to analyze electromyographic (EMG) signals, NNMF is a common and particularly suitable approach to isolate modules and their temporal activation profiles, and to decipher synergies during different functional tasks [3]. To perform NNMF, the electrical potentials recorded from several muscles during a task are first processed. Signal processing is used to take the raw signal data and transform it into a noise-free format that can be read by analysis algorithms. This is followed by filtering, data rectification, and smoothing of the EMG signal. The next step is NNMF, which allows for synergy identification. This method allows, from recordings of muscular activities obtained during a task, to identify several synergies by characterizing the relative activations of the muscle groups involved in a given synergy as well as the temporal activation profile of this synergy during the task. This method is very useful for characterizing the control and coordination of movements in the upper limbs (e.g.: reaching movements for an object) and in the lower limbs (e.g.: walking) [1][4].

Although many studies have investigated MS using NNMF [5][6][7], synergy extraction remains a complex process and is difficult to integrate into clinical practice due to its mathematical nature and to the lack of simple and intuitive tools for clinicians to integrate it into their assessment. To our knowledge, no existing tool automates the analysis of EMG signals through NNMF, posing a significant barrier to the advancement of knowledge in the field of movement and in the use of MS. In order to democratize the use of MS in clinical practice, our team have created a user-friendly tool to use in EMG analysis in order to make the extraction of MS using NNMF more accessible and intuitive for clinicians and researchers alike [11]. Our data collection process, as well as the mathematical principles underlying this tool and the use of NNMF for the extraction of MS are outlined below.

II. Methodology

A. Participants and Electrode Placement

Data was collected from 30 healthy adult participants between ages 18 and 35. The first step involved determining participants' preferred walking speed (PWS) on a split-belt treadmill. Surface electrodes (Cometa PICO EMG) were then installed on 8 muscles on the dominant leg, identified by [2] as the major components of walking synergies, in compliance with SENIAM standards (seniam.org): gluteus medius (GM), rectus femoris (RF), vastus medialis (VM), semitendinosus (ST), biceps femoris (BF), tibialis anterior (TA), medial gastrocnemius (MG) and soleus (SO). A flat sensor (footswitch) was also be placed under the calcaneus on the sole of the participant's shoe, allowing us to identify the start of a gait cycle.

B. Questionnaires

Participants were asked to fill out 2 validated questionnaires assessing pain-related beliefs: the Tampa Scale for Kinesiophobia (TSK) and the Pain Catastrophizing Scale (PCS). These questionnaires include statements pertaining to participants' pain-related beliefs (e.g.: rumination about pain; significance of pain) which participants had to rate their level of agreement with. For example, the TSK questionnaire includes this statement: "Pain always means I have injured my body." [9] Similarly, the PCS questionnaire includes statements such as: "When I'm in pain, I anxiously want the pain to go away." [10] These two questionnaires are then graded on a numerical scale, with pre-established benchmarks in the literature demarcating clinically significant catastrophizing and kinesiophobia scores.

C. Pain-free Condition

Participants will be asked to perform a series of walking tests, both at a speed of 1.3 m/s and at their preferred walking speed. Each of these trials will consist of 60 steps, representing 30 iterations of the full gait cycle. Each trial will be recorded and labeled accordingly for analysis in later steps.

D. Pain Condition

Following the pain-free trials, the participant will be asked to sit comfortably on a chair and elevate their dominant leg on another chair so that it is accessible to the experimenter. A 1% capsaicin cream will be applied to an area of 10cm x 3cm on the back of the dominant knee. This cream will have the effect of inducing superficial pain in the skin of the lower limb. While the cream is taking effect, the participant will be asked to rate their pain level at one-minute intervals on a numerical rating scale ranging from 0 to 10 (0 being "no pain", 10 being "the worst pain imaginable"). Once their pain level has stabilized at or above a rating of 4/10 (i.e., the same

Funding for this project has been granted by the Natural Sciences and Engineering Research Council of Canada (NSERC) as well as by INTER – Ingénierie de technologies interactives en réadaptation, an initiative by the Fonds de Recherche du Québec – Nature et technologie (FRQ-NT).

rating on 2 consecutive measurements), the participant will be asked to return to the treadmill. The participant will be asked to repeat the walking task, i.e. 30 walking cycles with pain at 1.3 m/s, then 30 walking cycles at their PWS. Each trial will be recorded.

III. Data Analysis

Our tool will work through the J-based Gamma model steps described in [2].

A. Data Processing

Firstly, raw EMG are filtered using a fourth-order Butterworth zero-lag bandpass filter, with cutoff frequencies at 20 and 400 Hz before a centered 250-ms sliding RMS window was used to generate continuous EMG envelopes for each muscle. Each gait cycle is then normalized using 101 temporal data points (0-100%) such that different cycles, regardless of duration, can be compared.

Following normalization, the preprocessed EMG signal (EMG_{exp}) will be decomposed into 2 matrices (W, H) using NNMF.

$$W' = m * s \quad \text{, where:} \tag{1}$$

W: the muscle weighting matrix (contribution of each muscle in each synergy)

m = muscle

s = synergy

$$H = s * t \quad \text{, where:} \tag{2}$$

H = temporal pattern (activation profile for each synergy during a cycle)

m = muscle

s = synergy

Each W matrix will then be normalized relative to its maximum value.

$$W = Weight * W' \tag{3}$$

These matrices will be multiplied create an approximated reconstruction of the original EMG signal (EMG_{rec}). The algorithm will then evaluate the Variance Accounted For (VAF) between the EMG_{exp} and the EMG_{rec}:

$$VAF = 1 - \left(\frac{\sum \left(EMG_{exp} - EMG_{rec} \right)^2}{\sum {EMG_{exp}}^2} \right) \tag{4}$$

Finally, it will determine of the minimum amount of muscle synergies so that the VAF of each muscle (VAF_m) is greater than 90% and the overall VAF (VAF_g) is greater than 80%. The algorithm will automatically stop when either (a) the absolute difference between coefficients is < 1e-8 for 20 consecutive iterations or (b) 500 iterations occur without any convergence. This process is repeated 20 times and the result with the least reconstruction error with the smallest number of synergies is kept. These settings can all be adjusted within the synergy analysis tool.

B. Statistical Analysis

Cosine Similarity on Center of Activity (CoA)

- The cosine similarities on the activity center represent the center of the activation distribution during the gait cycle and allow to compare synergies temporally. To do this, the distribution is expressed in the form of a circle using circular statistics, where 0° and 180° represent 0% and 50% of the gait cycle, respectively. We can thus define the activity center as the angle of the vector pointing towards the center of distribution of the activity in a cycle. The following equations allow to calculate it:

$$A = \sum_{t=1}^{N} (\cos \theta_t \times H_t) \tag{5}$$

$$B = \sum_{t=1}^{N} (\sin \theta_t \times H_t) \tag{6}$$

$$CoA = \tan^{-1}\left(\frac{B}{A}\right) \quad \text{, where :} \tag{7}$$

θ_t : the angle at point *t* of the normalized cycle (1-100) transforming the cycle (1-100%) into an angle (3,6-360°).

H_t : the activation amplitude of the synergy at point *t*.

We will measure intraclass correlation coefficients between pain-free trials to measure the test-retest reliability of MS in a pain-free condition, as well as between pain-free and painful conditions for all participants. We anticipate a clinically significant difference between conditions. We will also be investigating the interplay of **pain-related beliefs** and **pain behavior** using the TSK and the PCS. We anticipate a larger difference between pain-free and painful conditions in participants scoring highly on these questionnaires.

C. *Tool Outputs and Interface*

An important facet and major strength of this tool is its ease of use and accessibility. Rather than outputting results and comparisons between synergies in a purely numerical fashion, this tool produces clear visuals to represent synergy comparisons.

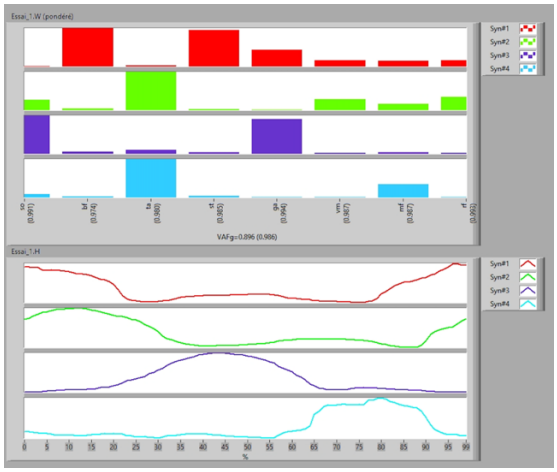


Fig. 1: Contributions of each muscle in different synergies during walking and temporal description of synergy occurrences through a gait cycle.

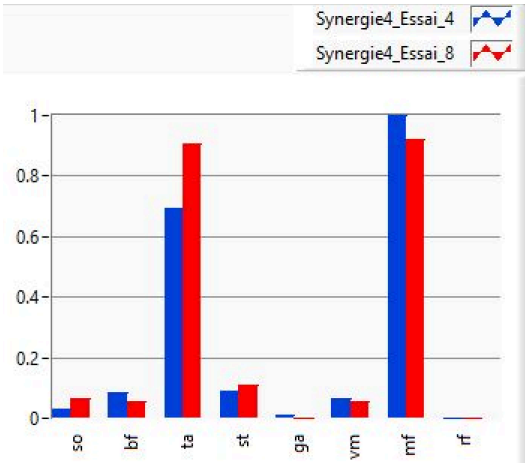


Fig. 2: Visual comparison of muscle synergies in pain-free conditions in two different trials.

IV. Clinical Significance

The present study will establish and document the test-retest of NNMF for the analysis of MS during human locomotion. In doing so, the results will contribute to improving the interpretation of previous work done with this measurement approach and to increasing the scientific rigor of future studies on MS. The present work will also allow us to better understand the influence of pain and psychological factors related to pain on MS, thus contributing to a better understanding of how these factors can affect gait and motor skills. This study will also establish a protocol to extract muscle synergies in a uniform manner using an automated tool, which will make this complex process much more accessible to clinicians and researchers.

Looking ahead, future studies can aim to implement MS analysis in diagnostic tests, for example in the case of stroke or Parkinson's disease (PD), where abnormal movement patterns may be present. In PD, for example, slight abnormalities in locomotion can exist even early in the disease, with a more pronounced presence on one side of the body [8]. Studies may seek to refine detection thresholds using MS analysis to be able to detect these minute changes earlier than those visible to the naked eye, which would allow an earlier diagnosis of this type of neurodegenerative disease as well as a concrete quantitative basis for the progression of the disease. By collecting longitudinal data from patients in a clinical setting in further studies, studies can aim to characterize pathology progression in this population. The use of MS is may also be particularly relevant in the analysis of motor function recovery, whether after a musculoskeletal injury or following a stroke. This study aims to create a basis off which future explorations of muscle synergies can be based.

REFERENCES

[1] S. Safavynia, G. Torres-Oviedo, and L. Ting, "Muscle Synergies: Implications for Clinical Evaluation and Rehabilitation of Movement," *Topics in Spinal Cord Injury Rehabilitation*, vol. 17, no. 1, pp. 16–24, Jul. 2011, doi: 10.1310/sci1701-16.

[2] M. J. Escalona, D. Bourbonnais, M. Goyette, D. Le Flem, C. Duclos, and D. H. Gagnon, "Effects of Varying Overground Walking Speeds on Lower-Extremity Muscle Synergies in Healthy Individuals," *Motor Control*, vol. 25, no. 2, pp. 234–251, Apr. 2021, doi: 10.1123/mc.2020-0008.

[3] M. F. Rabbi, C. Pizzolato, D. G. Lloyd, C. P. Carty, D. Devaprakash, and L. E. Diamond, "Non-negative matrix factorisation is the most appropriate method for extraction of muscle synergies in walking and running," *Sci Rep*, vol. 10, no. 1, p. 8266, May 2020, doi: 10.1038/s41598-020-65257-w.

[4] L. H. Ting and J. L. McKay, "Neuromechanics of muscle synergies for posture and movement," *Current Opinion in Neurobiology*, vol. 17, no. 6, pp. 622–628, Dec. 2007, doi: 10.1016/j.conb.2008.01.002.

[5] L. Hagedoorn, M. Zadravec, A. Olenšek, E. Van Asseldonk, and Z. Matjačić, "The Existence of Shared Muscle Synergies Underlying Perturbed and Unperturbed Gait Depends on Walking Speed," *Applied Sciences*, vol. 12, no. 4, p. 2135, Feb. 2022, doi: 10.3390/app12042135.

[6] B. Kibushi, S. Hagio, T. Moritani, and M. Kouzaki, "Speed-Dependent Modulation of Muscle Activity Based on Muscle Synergies during Treadmill Walking," *Front. Hum. Neurosci.*, vol. 12, p. 4, Jan. 2018, doi: 10.3389/fnhum.2018.00004.

[7] M. Kristiansen, A. Samani, P. Madeleine, and E. A. Hansen, "Muscle synergies during bench press are reliable across days," *Journal of Electromyography and Kinesiology*, vol.30, pp. 81-88, Oct. 2016, https://doi.org/10.1016/j.jelekin.2016.06.004

[8] J. C. Greenland, R. A. Barker, and John Van Geest Centre for Brain Repair, Department of Clinical Neurosciences, University of Cambridge, UK, "The Differential Diagnosis of Parkinson's Disease," in Parkinson's Disease: Pathogenesis and Clinical Aspects, Codon Publications, 2018, pp. 109–128. doi. 10.15586/codonpublications.parkinsonsdisease.2018.ch6.

[9] R. P. Miller, S. H. Kori, D. D. Todd, "The Tampa Scale: a Measure of Kinisophobia," *The Clinical Journal of Pain* 7(1):p 51, March 1991.

[10] M. J. L. Sullivan, S. R. Bishop, and J. Pivik, "The Pain Catastrophizing Scale: Development and validation," *Psychological Assessment, 7*(4), 524–532, 1995, https://doi.org/10.1037/1040-3590.7.4.524

[11] M.J. Escalona, D. Bourbonnais, M. Goyette, D. Le Flem, C. Duclos, and D. H. Gagnon, "Effects of Varying Overground Walking Speeds on Lower-Extremity Muscle Synergies in Healthy Individuals," *Motor control*, *25*(2), 234–251, 2021, https://doi.org/10.1123/mc.2020-0008

TagTrack: A Camera-Based System for Movement Tracking with Applications in Resistance Training

Sara Houde[1,3], Camille Martin[1,3], Eleonor Riesco[2,3], Renaud Tremblay[2,3], Karina Lebel[*,1,3]

[1]Faculty of Engineering, Université de Sherbrooke, Sherbrooke, QC, Canada, J1N 3C6
[2]Faculty of Physical Activity Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada, J1N 3C6
[3] Research Center on Aging, CIUSS de l'Estrie - CHUS, Sherbrooke, QC, Canada, J1H 2J7

[*]karina.lebel@usherbrooke.ca

***Abstract*— This study investigates the potential of TagTrack, a GoPro-based motion capture system using ArUco markers, to monitor neuromuscular fatigue during resistance exercise in clinical settings. Prescribing low-intensity resistance training poses challenges, especially in clinical environments where patients may struggle to accurately gauge effort using subjective measures. Therefore, there is a clear need for objective and accessible solutions. TagTrack, when affixed to various training machines, monitors the velocity profile during repeated sets and calculates a velocity loss index to objectively characterize effort and provide an indicator of neuromuscular fatigue.**

Keywords*— *Movement Signature, ArUco, Personalized Training, Pose Estimation

I. Introduction

The advent of portable cameras, such as GoPro and cellphone cameras, has transformed how we capture movement in sports and clinical settings. These systems offer a flexible, non-invasive way to monitor patients' movements and biomechanics, making them invaluable in different situations such as rehabilitation, physical therapy, and surgical training. The ability to track a patient's posture, gait, and muscle coordination with portable motion capture systems provides the potential for immediate clinical feedback on progress, enabling more effective adjustments to treatment. One area with significant potential for improvement through measurement systems is resistance training (RT). RT is widely recognized as the most effective intervention for increasing strength and muscle function in individuals with frailty [1]. In clinical and hospital settings, low loads are often considered due to challenges such as the fear of injury and limited access to specialized equipment [2] [3]. Low load training is defined as the use of low weights compared to the maximum weight the individual is able to lift for one repetition and expressed as a pourcentage (% 1-RM). However, for RT to remain effective at a low %1RM, it must be performed with a high level of perceived effort (*i.e.* high effort intensity), typically assessed subjectively using the individual's rate of perceived exertion (RPE) or repetitions-in-reserve (RIR) [4]. Traditional RT prescriptions, which assign a fixed %1RM with a corresponding number of repetitions, do not reliably reflect the level of effort exerted during a set. Accurately determining %1RM is difficult, if not impossible, in many clinical contexts, and the number of repetitions that can be performed at a given %1RM varies significantly based on several internal and external factors that are uncontrollable daily [5]. While RPE-based or RIR-based approaches attempt to address this issue, they also present limitations, as individuals may struggle to accurately assess their exertion, especially during high-repetition or low-intensity sets [6]. These challenges highlight the need for a more objective measure of effort in RT in clinical settings.

One approach to control those variabilities has been training to momentary muscle failure [7]. However, recent evidence suggests that similar adaptations (i.e., strength and hypertrophy) can be achieved when training close to, but not at, failure, thereby avoiding the significant negative impact of failure training on post-training recovery [8] [9]. Another promising approach involves monitoring movement velocity, as studies have suggested a close relationship between neuromuscular fatigue and movement velocity loss during resistance training [1] [10]. In other words, greater velocity loss during a set corresponds to higher effort intensity and being close to momentary muscle failure, leading researchers to explore velocity monitoring as a paradigm shift in RT prescription [11]. According to Hickmott *et al.* [5], a velocity loss level of $\geq 25\%$ has been deemed sufficient for beneficial adaptations. The most established method for velocity-based RT relies on linear position transducers such as GymAware (Kinetic Performance Technology, Canberra, Australia) [12] and T-Force dynamic measurement system (Ergotech Consulting, Murcia, Spain) [13]. While widely adopted in elite and amateur sports, these devices present several limitations in clinical settings. Their high cost (around 2000$ USD) and reliance on an Olympic barbell or strength machine restrict their use to specific exercises requiring a linear movement path [12] [14]. These constraints limit the accessibility of velocity-based RT in settings where alternative equipment, such as dumbbells or other types of strength training machines is more common. Given the benefits of velocity monitoring for RT prescription and the need for effort quantification at low %1RM, it is crucial to develop accessible solutions that can be used in various clinical and rehabilitation environments. In this context, camera-based motion capture presents promising opportunities.

Open source libraries, like OpenCV or MediaPipe, provide algorithms for body detection and pose estimation from images captured by Red Green Blue (RGB) cameras. These techniques are widely used in video games, augmented reality, navigation, and more. However, their reliability for clinical applications remains uncertain, as their robustness can be affected by factors such as body features, clothing, environment, and image quality [15] [16]. Another challenge is depth measurement, which is unavailable using a single RGB camera. Position coordinates are sometimes estimated based on the average dimensions of certain body features within a given population [17]. For example, the depth of a person can be estimated by knowing the standard distance between eyes. However, this approach can introduce significant errors. Some cameras include a depth components (RGB-D). However, most RGB-D devices are not stand-alone; they require a computer or embedded system to operate, thereby increasing the cost and complexity of the platform and reducing its portability. An alternative method for obtaining accurate depth and orientation information in RGB images involves the use of pattern recognition algorithms [18]. While pattern recognition algorithms can be trained to recognize specific objects known to be present within the scene, a popular approach considers the use of specific markers or patterns to facilitate the process. Among the various patterns proposed in the literature, the ArUco marker from the OpenCV library is the most popular one (Figure 1a) [19] [18]. ArUcos are binary square fiducial markers that serve as reliable reference points for cameras. These markers are made of black and white cells, which together form a pattern with a unique identifier stored in a dictionary [20]. The known dimensions and specificities of the pattern used allow accurate estimation of distance and orientation, from the camera's point of view. ArUco markers are well-known for their robustness in various conditions, including normal and difficult angles, partial obstruction, and varying lighting conditions [18] [21]. The primary advantages of this approach are accuracy and low cost [21]. Thus, combining a GoPro with an ArUco marker appears to be an accessible and cost-effective solution for precise movement tracking. This paper presents the TagTrack system, a straightforward and adaptable system designed to facilitate knowledge translation across different application domains. Its use in a specific RT context is then presented and discussed.

II. TagTrack: a GoPro-based Measurement System Utilizing ArUco Markers for movement Analysis

The proposed system is composed of an RGB camera, the GoPro Hero 9, and an ArUco marker. The GoPro Hero 9 camera offers several advantages, including portability, image stability, and affordability. The latest model, priced around 550$ CAD, is significantly more accessible than traditional motion capture systems [22] [23] [24]. The camera is recording at 60Hz, in linear mode with a 4K resolution. The ArUco markers are fetched directly from the OpenCV library (v. 4.5.3). The marker size was carefully selected to ensure accuracy, facilitate fixation, and be visible within a range of 0.5m to 4m, which is optimal for ArUco detection accuracy [21] and is within the typical range of training machines. A 5X5 pixel ArUco marker of 56 mm is fixed, face up, on the top of the training machine load. ArUco markers are usually simply printed patterns, which can present issues with contrast and alignment depending on printer quality. In TagTrack, the markers are constructed using black and white LEGO blocs, chosen for the high precision of LEGO's fabrication process. A matte varnish was applied to reduce the potential errors due to reflections of shiny surfaces. To maximize accuracy in the detection of the ArUco markers, the camera was calibrated using the OpenCV calibration procedure [25] [26].

A. TagTrack for Velocity-Based RT Monitoring: Adaptations and Potential Assessment

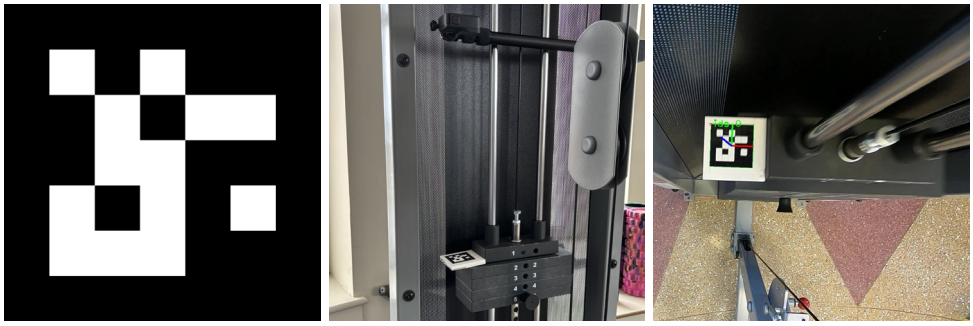
To use TagTrack in a RT context, specific positioning and adaptations are required. First, the camera is mounted on the side of the machine, near the top, using an adaptive bar mount (*e.g.*, Manfrotto). The GoPro is positioned facing down to capture the ArUco as it moves up and down along the depth axis of the image (Figure 1b). To assess the potential of the TagTrack system in RT context, an emulated resistance training session with a healthy participant performing a strength training program designed for older women with hypertension. This training includes multiple sets of 10 repetitions on the following equipment:

- Exercise 1: Seated Leg Press (Seated Leg Press Pro 2 Series, LifeFitness, CA)
- Exercise 2: Bench Press (Multipress Fit Series, LifeFitness, CA)
- Exercise 3: Seated Leg Curl (Seated Leg Curl Pro 2 Series, LifeFitness, CA)
- Exercise 4: Close-Grip Low Row (Lat Pulldow/Low Row Fit Series, LifeFitness, CA)

Camera recordings were initiated once per exercise, resulting in a single recording for the repeated sets. To facilitate interactions, recordings were triggered and ended through vocal commands. Information about the 3D position of the ArUco relative to the camera is extracted using the OpenCV library in Python. Velocity is computed from the position's norm derivation. Segmentation of sets of repetitions is achieved by searching for stable moments, which are computed using a sliding window technique. Peak velocities, per repetition, are then identified. The velocity loss index is finally calculated with a technic inspired by Sousa *et al.* [13]. Figure 2 presents the data analysis process on a single series of the Close-Grip Low Row. In this case, the mean propulsive velocity (MPV) of the fastest repetition is 1.44 m/s while the lowest is 0.60 m/s, leading to a velocity loss of 58%.

III. Discussion & Conclusion

This exploratory study demonstrates the potential of using TagTrack to assess neuromuscular fatigue during RT in clinical settings. In the near future, the TagTrack system will be used during RT training of women with hypertension as an exploratory



(a) 5x5 ArUco id 0 [27]. (b) ArUco on the machine. (c) Detection from the camera.

Fig. 1: ArUco and its system application.

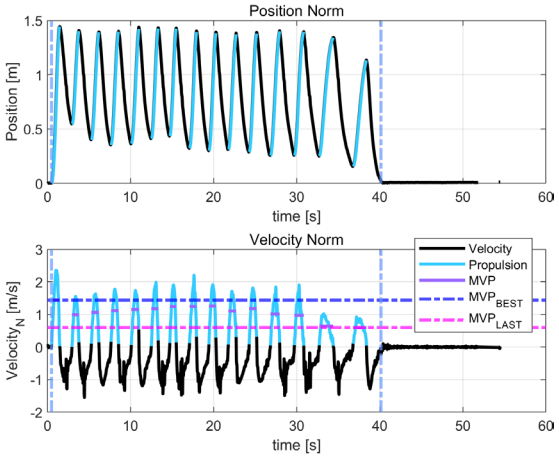


Fig. 2: TagTrack Velocity Profile Processing.



Fig. 3: Future applications for TagTrack.

measurement, to be compared with conventional subjective measures. To enhance the system's accuracy, improvements can be made to the physical setup. For instance, Figure 1c shows motion blur caused by an object in movement [28]. This fast motion can introduce errors in the ArUco estimation, which can be reduced by aligning the camera to the motion axis. To minimize this phenomenon, a rigid bracket will be designed to hold the ArUco marker perpendicular to the camera. Additionally, designing a custom camera bar mount to keep the camera perpendicular to the side panel of the machine could improve accuracy. Another consideration is to reverse the position of the ArUco and the GoPro to make use of the image stabilizer embedded within the camera. It is hypothesized that moving the camera instead of the ArUco would improve image quality, given the recognized efficiency of the GoPro image stabilizer. This proof of concept opens the door to using TagTrack on various RT devices beyond linear training machines. A similar analysis process could be implemented to track more complex motions, such as placing markers on free weights or dumbbells (Figure 3). Following the planned RT study, the development of a mobile application capable of running the algorithm in real-time will be discussed with the exercise specialist researchers. This would provide clinicians with immediate access to information during the test, enabling them to halt repetitions at the optimal stage of fatigue and thus better individualize training. Finally, movement analysis frequently focuses on the speed of body segments, particularly during training, as it is closely linked to muscular responses. To further this concept, several studies have introduced the Sigma-Lognormal model, which establishes a direct connection between neuromuscular response and movement speed by extracting relevant parameters. The velocity profile provides insights into a group of muscles and explains how movements are generated and synchronized [29]. This type of analysis could be used to enhance the analysis of neuromuscular fatigue. In conclusion, the TagTrack system shows promise as a cost-effective and accessible tool for monitoring neuromuscular fatigue in clinical settings. By addressing the current limitations and exploring further applications, TagTrack could significantly improve the precision and effectiveness of resistance training programs, particularly for populations with specific clinical needs.

Acknowledgment

This manuscript was reviewed for language and fluidity with the assistance of Microsoft Copilot. This study is partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada Research Chair in Biomechanical Movement Signature to Better Understand and Reduce Disability, and the Research Chair for Healthy Aging.

References

[1] P. Lopez, R. S. Pinto, R. Radaelli, A. Rech, R. Grazioli, M. Izquierdo, and E. L. Cadore, "Benefits of resistance training in physically frail elderly: a systematic review," *Aging Clinical and Experimental Research*, vol. 30, no. 8, pp. 889–899, Aug. 2018. [Online]. Available: https://doi.org/10.1007/s40520-017-0863-z

[2] E. Van Roie, I. Bautmans, W. Coudyzer, F. Boen, and C. Delecluse, "Low- and High-Resistance Exercise: Long-Term Adherence and Motivation among Older Adults," *Gerontology*, vol. 61, no. 6, pp. 551–560, May 2015, _eprint: https://karger.com/ger/article-pdf/61/6/551/2837589/000381473.pdf.[Online]. Available: https://doi.org/10.1159/000381473

[3] F. Dijkstra, G. van der Sluis, H. Jager-Wittenaar, L. Hempenius, J. S. M. Hobbelen, and E. Finnema, "Facilitators and barriers to enhancing physical activity in older patients during acute hospital stay: a systematic review," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 19, no. 1, p. 99, Jul. 2022. [Online]. Available: https://doi.org/10.1186/s12966-022-01330-z

[4] J. Jansen, P. W. Marshall, J. R. Benatar, R. Cross, T. K. Lindbom, and M. Kingsley, "Low-Intensity Resistance Exercise in Cardiac Rehabilitation: A Narrative Review of Mechanistic Evidence and Clinical Implications," *Journal of Clinical Medicine*, vol. 13, no. 23, p. 7338, Dec. 2024. [Online]. Available: https://www.mdpi.com/2077-0383/13/23/7338

[5] L. M. Hickmott, P. D. Chilibeck, K. A. Shaw, and S. J. Butcher, "The Effect of Load and Volume Autoregulation on Muscular Strength and Hypertrophy: A Systematic Review and Meta-Analysis." *Sports medicine - open*, vol. 8, no. 1, p. 9, Jan. 2022.

[6] I. Halperin, T. Malleron, I. Har-Nir, P. Androulakis-Korakakis, M. Wolf, J. Fisher, and J. Steele, "Accuracy in Predicting Repetitions to Task Failure in Resistance Exercise: A Scoping Review and Exploratory Meta-analysis," *Sports Medicine*, vol. 52, no. 2, pp. 377–390, Feb. 2022. [Online]. Available: https://link.springer.com/10.1007/s40279-021-01559-x

[7] S. J. Dankel, M. B. Jessee, K. T. Mattocks, J. G. Mouser, B. R. Counts, S. L. Buckner, and J. P. Loenneke, "Training to Fatigue: The Answer for Standardization When Assessing Muscle Hypertrophy?" *Sports Medicine*, vol. 47, no. 6, pp. 1021–1027, Jun. 2017. [Online]. Available: http://link.springer.com/10.1007/s40279-016-0633-7

[8] R. Morán-Navarro, C. E. Pérez, R. Mora-Rodríguez, E. de la Cruz-Sánchez, J. J. González-Badillo, L. Sánchez-Medina, and J. G. Pallarés. "Time course of recovery following resistance training leading or not to failure," *European Journal of Applied Physiology,* vol. 117, no. 12, pp. 2387-2399, Dec. 2017. https://doi.org/10.1007/s00421-017-3725-7.

[9] A. F. Vieira, D. Umpierre, J. L. Teodoro, S. C. Lisboa, B. M. Baroni, M. Izquierdo, and E. L. Cadore, "Effects of Resistance Training Performed to Failure or Not to Failure on Muscle Strength, Hypertrophy, and Power Output: A Systematic Review With Meta-Analysis," *Journal of Strength and Conditioning Research*, vol. 35, no. 4, pp. 1165–1175, Apr. 2021. [Online]. Available: https://journals.lww.com/10.1519/JSC.0000000000003936

[10] K. Shattock and J. C. Tee, "Autoregulation in Resistance Training: A Comparison of Subjective Versus Objective Methods," *Journal of Strength and Conditioning Research*, vol. 36, no. 3, pp. 641–648, Mar. 2022. [Online]. Available: https://journals.lww.com/10.1519/JSC.0000000000003530

[11] J. González-Badillo, L. Sánchez-Medina, J. Ribas-Serna, and D. Rodriguez Rosell. "Toward a New Paradigm in Resistance Training by Means of Velocity Monitoring: A Critical and Challenging Narrative." *Sports Medicine* - Open. 2022. [Online]. Available: doi:10.1186/s40798-022-00513-z

[12] S. N. Guppy, K. L. Kendall, and G. G. Haff, "Velocity-Based Training—A Critical Review," *Strength & Conditioning Journal*, vol. 46, no. 3, pp. 295–307, Jun. 2024. [Online]. Available: https://journals.lww.com/10.1519/SSC.0000000000000806

[13] T. Sousa, A. Pereira, D. L. Marques, H. P. Neiva, D. A. Marinho, and M. C. Marques, "Validity and reliability of the T-Force and Chronojump systems to measure movement velocity on resistance machines with older adults," *Scientific Reports*, vol. 14, no. 1, p. 23727, Oct. 2024. [Online]. Available: https://www.nature.com/articles/s41598-024-74159-0

[14] B. B. Appleby, H. Banyard, S. J. Cormack, and R. U. Newton, "Validity and Reliability of Methods to Determine Barbell Displacement in Heavy Back Squats: Implications for Velocity-Based Training," *Journal of Strength and Conditioning Research*, vol. 34, no. 11, pp. 3118–3123, Nov. 2020. [Online]. Available: https://journals.lww.com/10.1519/JSC.0000000000002803

[15] S. A. S. Mohamed, M.-H. Haghbayan, T. Westerlund, J. Heikkonen, H. Tenhunen, and J. Plosila, "A Survey on Odometry for Autonomous Navigation Systems," *IEEE Access*, vol. 7, pp. 97 466–97 486, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8764393/

[16] Y. Chen, Y. Tian, and M. He, "Monocular human pose estimation: A survey of deep learning-based methods," *Computer Vision and Image Understanding*, vol. 192, p. 102897, Mar. 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1077314219301778

[17] S. Thierfelder, "Create your own depth measuring tool with MediaPipe FaceMesh in Javascript." [Online]. Available: https://medium.com/@susanne. thierfelder/create-your-own-depth-measuring-tool-with-mediapipe-facemesh-in-javascript-ae90abae2362

[18] M. Kalaitzakis, B. Cain, S. Carroll, A. Ambrosi, C. Whitehead, and N. Vitzilaios, "Fiducial Markers for Pose Estimation: Overview, Applications and Experimental Comparison of the ARTag, AprilTag, ArUco and STag Markers," *Journal of Intelligent & Robotic Systems*, vol. 101, no. 4, p. 71, Apr. 2021. [Online]. Available: https://link.springer.com/10.1007/s10846-020-01307-9

[19] H. Sarmadi, R. Munoz-Salinas, M. A. Olivares-Mendez, and R. Medina-Carnicer, "Detection of Binary Square Fiducial Markers Using an Event Camera," *IEEE Access*, vol. 9, pp. 27 813–27 826, 2021. [Online]. Available:https://ieeexplore.ieee.org/document/9351958

[20] "Detection of ArUco Markers." [Online]. Available: https://docs.opencv.org/4.x/d5/dae/tutorial_aruco_detection.html

[21] B. S. R. Neto, T. D. O. Araújo, B. S. Meiguins, and C. G. R. Santos, "Tape-Shaped, Multiscale, and Continuous-Readable Fiducial Marker for Indoor Navigation and Localization Systems," *Sensors*, vol. 24, no. 14, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/24/14/4605

[22] "Hero 13 Black." [Online]. Available: https://gopro.com/fr/ca/shop/cameras/learn/hero13black/CHDHX-131-master.html

[23] E. Washabaugh, T. Kalyanaraman, P. Adamczyk, E. Claflin, and C. Krishnan, "Validity and repeatability of inertial measurement units for measuring gait parameters," *Gait and Posture*, vol. 55, pp. 87–93, 2017. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85018494043& doi=10.1016%2fj.gaitpost.2017.04.013&partnerID=40&md5=0072d8871ac6e34e0b3dfb42f61a2d03

[24] L. Ruiz-Ruiz, A. Jimenez, G. Garcia-Villamil, and F. Seco, "Detecting fall risk and frailty in elders with inertial motion sensors: A survey of significant gait parameters," *Sensors*, vol. 21, no. 20, 2021. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85117247001& doi=10.3390%2fs21206918&partnerID=40&md5=82677eb76c1838a4fda5f9b26726ba86

[25] "Camera Calibration." [Online]. Available: https://docs.opencv.org/4.x/dc/dbb/tutorial_py_calibration.html

[26] S. Kaustubh and M. Satya, "Camera Calibration using OpenCV," 2020. [Online]. Available: https://learnopencv.com/camera-calibration-using-opencv/

[27] K. Oleg, "ArUco markers generator!" [Online]. Available: https://chev.me/arucogen/

[28] "Comprendre les vitesses d'obturation lentes et rapides." [Online]. Available: https://www.adobe.com/ca_fr/creativecloud/photography/discover/shutter-speed.html

[29] C. Carmona-Duarte, M. A. Ferrer, R. Plamondon, A. Gómez-Rodellar, and P. Gómez-Vilda, "Sigma-Lognormal Modeling of Speech," *Cognitive Computation*, vol. 13, no. 2, pp. 488–503, Mar. 2021. [Online]. Available: https://doi.org/10.1007/s12559-020-09803-8

Temporal Memory Index: A Novel Feature for Handwriting Analysis in Developmental Dysgraphia

Ján Mucha*‡, Jiri Mekyska*†, Katarina Zvoncakova†,
*Department Af Telecommunications, Brno University Af Technology, Brno, Azech Republic
Email: mucha@vut.cz

†Department of Research Methodology, Institute of Psychology, Czech Academy of Sciences, Brno, Czech Republic
‡Tecnocampus Universitat Pompeu Fabra, Mataró, Spain

***Abstract*— Digital handwriting analysis offers a promising, non-invasive approach to assessing neurodevelopmental disorders like developmental dysgraphia (DD). Existing methods, while helpful, may not fully capture the subtle temporal dynamics and memory effects present in handwriting. This study introduces the Temporal Memory Index (TMI), a novel metric based on fractional order derivatives (FDs), designed to quantify these temporal dependencies. Using data from a fast Archimedean spiral task, we evaluated TMI's ability to differentiate between children with DD (n=155) and intact children (n=114). We extracted traditional kinematic features, FD features, and the proposed TMI features. While statistical tests showed no significant individual feature differences between groups, classification analysis (XGBoost) resulted in an increase in the specificity. Although modest in terms of overall classification accuracy, these preliminary findings suggest that TMI captures information about temporal dynamics in handwriting and warrants further investigation as a potential marker for DD.**

Keywords*— *temporal memory index, fractional calculus, developmental dysgraphia, graphomotor disabilities, online handwriting,

I. Introduction

Handwriting is a complex motor skill that reflects underlying neurological processes, making its analysis a valuable tool for assessing neurodevelopmental disorders. Digital handwriting analysis, in particular, offers a non-invasive and objective means of capturing subtle motor control deficits that might not be apparent through traditional observation [1].

Developmental dysgraphia (DD), a specific learning disability affecting handwriting fluency and legibility, presents a significant challenge for affected children [2]. Early and accurate identification of DD is crucial for implementing appropriate interventions. While traditional kinematic features (e.g., velocity, acceleration) and, more recently, fractional derivatives (FDs) have been used to analyze handwriting in DD [3], [4], these approaches may not fully capture the subtle temporal dynamics and memory effects inherent in the handwriting process. These temporal dependencies, reflecting the influence of past movements on current motor output, could be key to differentiating DD from typical handwriting. This limitation highlights the need for new analytical approaches that explicitly consider the temporal dependencies within handwriting movements.

This study introduces the Temporal Memory Index (TMI), a novel metric based on the difference between fractional derivatives of different orders, to quantify the degree of temporal memory in handwriting signals. Intuitively, TMI reflects the influence of past movements on the current state of a writer's pen trajectory, capturing information not readily available from instantaneous kinematic measures or even single-order FDs. We hypothesize that TMI could provide a more sensitive marker for distinguishing children with DD from intact children, potentially revealing subtle motor control differences not apparent with existing methods.

A. Temporal Memory Index (TMI) by Fractional Order Derivatives

II. Materials and Methods

Fractional calculus (FC) extends the concept of differentiation and integration to non-integer orders, providing a powerful tool for modeling systems with memory and long-range dependence [5]. Unlike integer-order derivatives, which are local operators, fractional derivatives incorporate information from the history of a signal. Our previous work has demonstrated the ability of FDs to enhance the differentiation of handwriting characteristics in neurodegenerative disorders [4] and children with DD [3], [6], suggesting their sensitivity to subtle nuances in motor control and coordination.

The Caputo fractional derivative is a commonly used definition appropriate for initial value problems. It is defined as [7]:

$$ {}^{C}\mathrm{D}^{\alpha} y(t) = \frac{1}{\Gamma(n-\alpha)} \int_{0}^{t} (t-\tau)^{n-\alpha-1} y^{n}(\tau)\, \mathrm{d}t, \tag{1} $$

where ${}^{C}\mathrm{D}^{\alpha} y(t)$ denotes the Caputo derivatives of order α of the function $y(t)$, Γ is the gamma function, and $n-1 < \alpha \leq n, n \in \mathbf{N}, t > 0$.

We propose a novel temporal memory index (TMI) based on the difference between fractional derivatives calculated with different orders to quantify the degree of temporal memory in handwriting signals. To compute TMI, we first define the alpha-difference vector (ADV) as:

$$ \mathrm{ADV(t)} = \mathrm{D}^{\alpha_2} f(t) - \mathrm{D}^{\alpha_1} f(t), \tag{2} $$

where $\alpha 2 > \alpha 1$. The ADV captures the difference in past influence on the signal, as represented by the two fractional derivatives. A larger difference indicates a stronger memory effect.

The TMI is then defined as the mean absolute value of the ADV over a given time interval:

$$\text{TMI} = \frac{1}{T}\int_0^T |\text{ADV(t)}|\text{d}t \approx \frac{1}{N}\sum_{i=1}^{N} |\text{ADV}(t_i)|, \tag{3}$$

where T is the time interval length, and N is the number of the samples.

To validate the proposed TMI and guide the selection of appropriate FDs (α_1, α_2) for the handwriting analysis, we conducted preliminary experiments using fractional Brownian motion (fBm) signals [8]. The fBm provides a well-established model for processes with controlled long-range dependence, characterized by the Hurst exponent (H), where $0 < H < 1$. Values of $H < 0.5$ indicate anti-persistence (short-term memory), $H = 0.5$ corresponds to standard Brownian motion (no long-range memory), and $H > 0.5$ indicates persistence (long-term memory). We generated fBm signals using the Davies-Harte method with three representative Hurst exponents: $H = 0.1$ (anti-persistent), $H = 0.5$ (no long-range memory), and $H = 0.8$ (persistent). For each H value, we calculated the TMI using three (α_1, α_2) pairs: (0.1, 0.5), (0.1, 0.9), and (0.5, 0.9). These pairs were chosen to represent varying degrees of difference between the FDs, allowing us to assess the TMI's sensitivity to this difference. The signals' lengths were set to 1000, with dt set to 0.01.

Figure 1 shows the calculated TMI values for the different H and (α_1, α_2) combinations. As expected, the TMI is lowest for $H = 0.5$, corresponding to standard Brownian motion with no long-range dependence. The TMI is significantly higher for $H = 0.1$, reflecting the anti-persistent nature of the signal. The large difference between α_1 and α_2 captures the discrepancy between short-term fluctuations and a longer-term "memory" that is not truly present. The TMI is also elevated for $H = 0.8$ compared to $H = 0.5$ but lower than for $H = 0.1$. This indicates that while the long-range dependence influences both fractional derivatives, the smaller α_1 still captures more short-term variations, leading to a non-zero difference. Furthermore, for a fixed H, the TMI is consistently higher for the (0.1, 0.9) pair, representing the largest difference in memory lengths considered. This highlights the influence of the past on the signal.

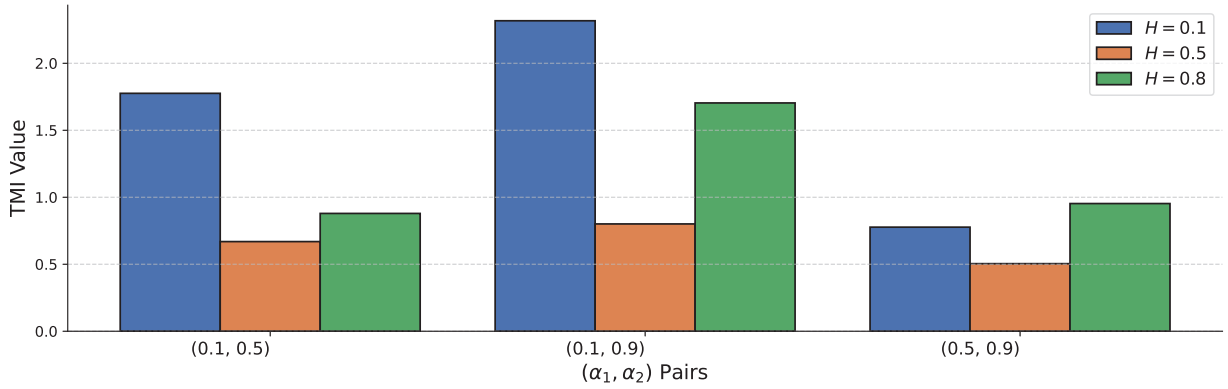


Fig. 1. Temporal Memory Index (TMI) values for different Hurst exponent (H) and (α_1, α_2) pairs.

B. Database

We enrolled 269 children from Czech primary schools, stratified into two groups: intact children (IC, 114) and those with developmental dysgraphia (DD, 155), as diagnosed by a special education needs teacher. Children were instructed to complete a fast Archimedean spiral on a Wacom Cintiq 16 (DTK1660K0B) display tablet during data acquisition. The fast spiral task was chosen for this exploratory study to examine potential differences in motor control between the groups. Informed consent was obtained from all parents, and the study was approved by the Ethics Committee of the Institute of Psychology at the Czech Academy of Sciences, adhering to the American Psychological Association's Ethical Principles [9].

C. Feature Extraction

We analyzed only the on-surface movement and computed the basic kinematic features, velocity, acceleration, and jerk, along with their horizontal (x) and vertical (y) components. To capture temporal memory characteristics, we then calculated FDs of the same kinematic features using the Caputo definition (Equation 1) and the orders $\alpha = \{0.1, 0.3, 0.5, 0.7, 0.9\}$. These values were chosen to span a range of memory lengths, from short-term to longer-term influences.

We computed the novelly proposed TMI using all pairwise combinations of those above α values. This allowed us to explore the sensitivity of TMI to different memory length comparisons. All features were calculated per stroke, and the mean,

standard deviation, and median were computed to get scalar values, yielding 18 baseline kinematic features, 90 FD features, and 180 TMI features (288 total).

D. Statistical Analysis and Machine Learning

Prior to statistical analysis, we considered grade and gender as confounding factors and removed their effects using linear regression [10]. Subsequently, we calculated Spearman's correlation (with False Discovery Rate (FDR) correction for multiple comparisons) to explore the relationships between handwriting features and clinical group (IC/DD). We assessed the normality of each feature distribution using the Shapiro-Wilk test to compare feature distributions between the DD and IC groups, and we employed t-tests (significance threshold: adjusted p-value < 0.05).

Finally, we performed a classification analysis using the XGBoost algorithm to evaluate the discriminative power of the features. We used a 10-fold cross-validation with five repetitions. With Matthew's Correlation Coefficient (MCC) as the primary optimization metric, a Bayesian optimization strategy was employed to fine-tune the hyperparameters. The model's performance was evaluated using balanced accuracy (BACC), sensitivity (SEN), specificity (SPE), and MCC. We assessed feature importance to quantify each feature's contribution to the classification.

III. Results

Based on the results of correlation analysis, various features derived from the TMI showed a positive correlation with the clinical status. However, after adjusting the p-values, only one TMI feature (median TMI of velocity (α_1=0.5, α_2=0.7) exhibited a borderline significant correlation ($p = 0.058$). In contrast, nine FD-based features derived from the velocity (predominantly horizontal velocity) showed statistically significant positive correlations with the clinical status.

Following the correlation analysis, we conducted t-tests between the IC and DD groups. After the p-value adjustment, none of the individual features showed a statistically significant difference between the two groups.

TABLE I. Results of the Classification Performance

Feature Set	Balanced Accuracy	Sensitivity	Specificity	MCC
All features (baseline + FD + TMI)	0.579 ± 0.070	0.854 ± 0.099	0.304 ± 0.120	0.201 ± 0.177
Baseline + FD (without TMI)	0.580 ± 0.077	0.717 ± 0.125	0.442 ± 0.115	0.171 ± 0.163

[1] FD – Fractional order derivative; TMI – Temporal Memory Index.

Table I presents the classification results using XGBoost. Classification performance was modest for both feature sets. BACC was slightly above chance (0.5) for both all features (0.579 ± 0.070) and baseline + FD features (0.580 ± 0.077). Excluding TMI features resulted in an increase in specificity (from 0.304 ± 0.120 to 0.442 ± 0.115) and a decrease in sensitivity (from 0.854 ± 0.099 to 0.717 ± 0.125).

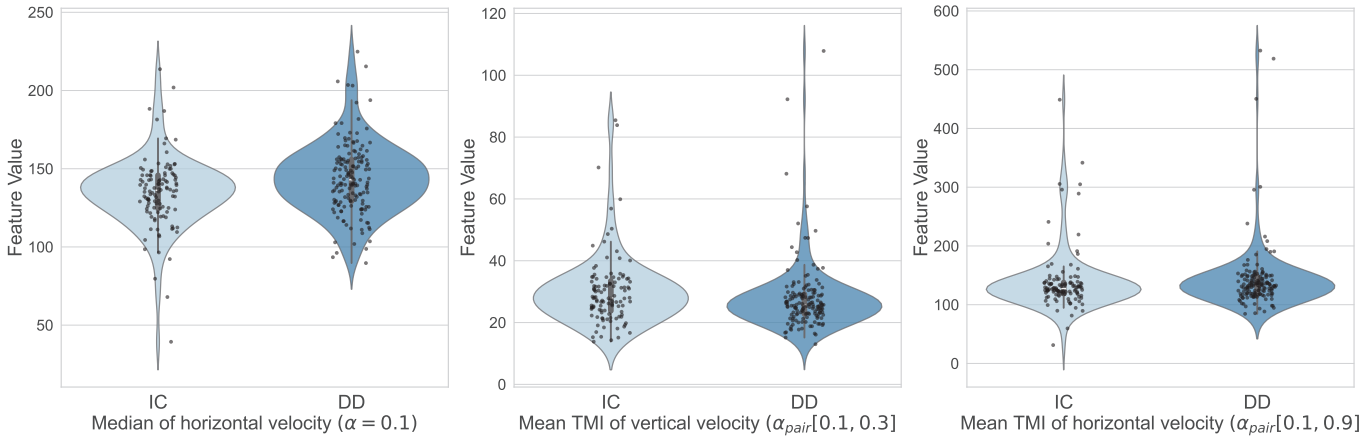


Fig. 2. Distribution of selected features across IC and DD groups.

Figure 2 illustrates the distributions of three representative features: a non-TMI feature (median horizontal velocity) and two TMI features (mean TMI of vertical and horizontal velocities). These distributions are examined to provide a visual context for the statistical and classification results.

IV. Discussion and Conclusion

This study introduced a novel Temporal Memory Index (TMI) based on FDs to quantify temporal memory effects in handwriting signals, specifically in the context of differentiating children with DD from IC. Our primary goal was to explore the potential of TMI as a sensitive marker for subtle motor control differences associated with DD.

The initial correlation analysis and t-test findings did not uncover exceptionally significant results. Nevertheless, this aligns with our previous research highlighting the importance of advanced and more robust parameters [3]. The inherent variability within the DD group and the relatively small sample size in each group after stratification likely contributed to the lack of statistically significant differences at the individual feature level [1]. However, it is crucial to recognize that our results are preliminary and require further validation with larger, more diverse cohorts. Moreover, the performance of TMI is comparable to that of the FD and could deserve further investigation.

Figure 2 presents the distributions of representative TMI and non-TMI features, visually demonstrating the complexity of differentiating between IC and DD groups based on single features. The considerable overlap in distributions across all features underscores the multifaceted nature of dysgraphia and the need for multifaceted analytical approaches.

The classification analysis, using XGBoost, further supports this notion. While the overall classification performance was modest, including TMI features, it did not substantially degrade performance compared to only baseline and FD features. The decrease in sensitivity when TMI features were excluded, coupled with an increase in specificity, suggests that TMI might contribute to a more nuanced understanding of the underlying differences, even if it doesn't dramatically improve overall classification accuracy. This pattern—modest overall performance but hints of differential contributions from different feature sets is typical in early-stage research exploring complex neurological conditions [6]. The temporal memory effect, as captured by TMI, might reflect subtle timing and coordination deficits that are not fully captured by traditional kinematic or even FD-based features.

This study has several limitations. First, we focused on a single handwriting task (fast Archimedean spiral). Future research should explore the performance of TMI across a broader range of handwriting tasks, varying in complexity and motor demands, to better understand its generalizability. Second, our TMI calculation was relatively simple, based on the mean absolute difference between FDs. More sophisticated approaches, potentially incorporating weighting schemes or different distance metrics, could enhance the sensitivity of TMI. Third, this study focused exclusively on DD. Investigating the behavior of TMI in other neurodevelopmental and neurodegenerative disorders characterized by motor control deficits (e.g., Parkinson's disease) is crucial to determine its broader applicability.

In conclusion, this work introduces the Temporal Memory Index as a potentially valuable tool for analyzing temporal memory effects in handwriting signals.

Acknowledgment

This work was supported by the Czech Science Foundation, project no. GN23-06074O (Research of advanced methods of graphomotor disabilities analysis based on fractional calculus), by the project Mediated Society (MEDIS:ON) CZ.02.01.01/00/23_025/0008713 co-financed/supported by the European Union and Czech Republic, and project no. 25-15668S supported by the Czech Science Foundation.

References

[1] C. Lopez and L. Vaivre-Douret, "Concurrent and predictive validity of a cycloid loops copy task to assess handwriting disorders in children," *Children*, vol. 10, no. 2, p. 305, 2023.

[2] P. J. Chung, D. R. Patel, and I. Nizami, "Disorder of written expression and dysgraphia: definition, diagnosis, and management," *Translational pediatrics*, vol. 9, no. Suppl 1, p. S46, 2020.

[3] Z. Galaz, J. Mucha, V. Zvoncak, J. Mekyska, Z. Smekal, K. Safarova, A. Ondrackova, T. Urbanek, J. M. Havigerova, J. Bednarova *et al.*, "Advanced parametrization of graphomotor difficulties in school-aged children," *IEEE Access*, vol. 8, pp. 112 883–112 897, 2020.

[4] J. Mucha, M. Gavenciak, J. Mekyska, M. Faundez-Zanuy, L. Brabenec, and I. Rektorova, "Evaluation of fractional calculus and delta parameters in prodromal diagnosis of dementia with Lewy bodies utilizing online handwriting," in *2024 32nd European Signal Processing Conference (EUSIPCO)*, 2024, pp. 1731–1735.

[5] I. Podlubny, *Fractional differential equations an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications*. San Diego: Academic Press, 1999.

[6] M. Gavenciak, J. Mucha, J. Mekyska, Z. Galaz, K. Zvoncakova, and M. Faundez-Zanuy, "Computer-aided diagnosis of graphomotor difficulties utilizing direction-based fractional order derivatives," *Cognitive Computation*, vol. 17, no. 1, p. 13, 2025.

[7] M. Caputo, "Linear Models of Dissipation whose Q is almost Frequency Independent—II," *Geophysical Journal International*, vol. 13, no. 5, pp. 529–539, 1967.

[8] B. B. Mandelbrot and J. W. Van Ness, "Fractional brownian motions, fractional noises and applications," *SIAM review*, vol. 10, no. 4, pp. 422–437, 1968.

[9] A. P. Association *et al.*, "Ethical principles of psychologists and code of conduct," 2016.

[10] M. T. Todd, L. E. Nystrom, and J. D. Cohen, "Confounds in multivariate pattern analysis: theory and rule representation case study," *Neuroimage*, vol. 77, pp. 157–165, 2013.



Session

Neuroscience & Applications

An Examination of Handwriting Inclination and Size as an Indicator of Emotional States: Separating Fact from Pseudoscience

Marcos Faùndez-Zanuy*, Anna Esposito†, Gennaro Cordasco†
*Tecnocampus, Universitat Pompeu Fabra, Spain
faundez@tecnocampus.cat
†Department of Psychology, Second University of Naples, Italy
{anna.esposito, gennaro.cordasco}@unina2.it

Keywords: handwritten text; graphology; inclination; DASS.

Abstract- This paper evaluates the relationship between handwriting inclination, x-height, and emotional states such as depression, anxiety, and stress. Using the EMOTHAW database, which includes handwriting samples from 129 participants along with Depression, Anxiety and Stress Scale (DASS) emotional state assessments, we extracted key handwriting features: inclination angle and x-height. Experimental results reveal no significant correlation between inclination and body text height with depression and stress. These findings provide further evidence against the scientific validity of certain graphology claims related to personality. However, we found a small significant correlation between height and anxiety.

I. Introduction

Handwriting has long been viewed as a window into the mind, with graphology historically claiming to reveal personality traits through the analysis of script. However, such claims have largely been dismissed by the scientific community due to a lack of empirical support. In contrast, contemporary psychology and behavioral science are beginning to explore handwriting not as a diagnostic tool for personality, but as a motor behavior potentially influenced by emotional states. This study follows this empirical approach, investigating whether measurable handwriting features—specifically slant and size—are associated with levels of depression, anxiety, and stress in a young adult population. The Depression scale has subscales assessing dysphoria, hopelessness, devaluation of life, self-deprecation, lack of interest/involvement, anhedonia, and inertia. The Anxiety scale assesses autonomic arousal, skeletal muscle effects, situational anxiety, and subjective experience of anxious affect. The Stress scale's subscales highlight levels of non-chronic arousal through difficulty relaxing, nervous arousal, and being easily upset/agitated, irritable/over-reactive, and impatient.

A. Graphology versus psychology

The oldest and widespread method to identify and understand certain psychological characteristics, such as personality through handwriting analysis is graphology [1]. The art of graphology has always fascinated and interested the public; the first historical notes on graphology can be attributed to various philosophers such as Aristotle and Suetonius, the latter analyzing the handwriting of Emperor Augustus. The first text dealing with graphology was written in 1622 by philosopher Camillo Baldi in his treatise "Come da una lettera missiva si conoscano la natura e qualità dello scrittore" (*Just as from a letter, one can discern the nature and qualities of the writer.*). The term graphology was coined, around 1871, by Jean Hippolyte Michon.

The observable graphological features are somewhat unique to the writer in that they represent the functioning of the neuroanatomical system. This system, involved in writing, is particularly complex, as it relies on both functions belonging to the pyramidal system (voluntary motility) and the extrapyramidal system (automatic and associated motility) [2]. A fundamental aspect of graphology is that of spatial symbolism, introduced by Max Pulver, characterized by spatial dimensions and vectors [3]. Nowadays, graphology and handwriting analysis can be used in different fields such as person analysis [4], career guidance [5] and neurodegenerative diseases [6]–[8].

For instance, studies [9] and [10] adopted a graphological analysis approach performed by forensic experts, aiming to identify features in the graphic content specific to depression (e.g., tremors in writing, descending strokes, irregularly sized letters, letter shapes). However, these studies reported no statistically significant differences between controls and patients with depression.

Unfortunately, graphology is based on subjective nonstandardized evaluations carried out by graphologists requiring a very costly and time-consuming analyses.

B. Graphology

In graphology [11], [12], [13], [14] (the study of handwriting to analyze personality traits), the analysis of mood disorders such as depression or euphoria involves examining various characteristics of the handwriting. The tendency of the line—whether it moves upward or downward—is one key indicator of the writer's emotional state, as summarized next:

1) Depression: Downward Slant of the Line: A key feature in identifying depression through handwriting is when the writing shows a downward slant, especially towards the end of the line or on the letters. This could suggest a lack of motivation, energy, or a sense of despair. The downward trend may represent a person's tendency to feel down or deflated.

Other Indicators include:

- **Pressure**: Excessive pressure on the paper may indicate inner tension or stress, often linked to depression.
- **Small Writing**: Small, cramped writing may suggest introversion or withdrawal, often associated with depression.
- **Unsteady or Shaky Lines**: This can reflect emotional instability, a characteristic often seen in people experiencing depression.
- **Tightness**: Tight, controlled strokes may indicate anxiety or emotional suppression, which often accompanies depression.

2) Euphoria: Upward Slant of the Line: On the other hand, a handwriting that tends to slope upwards, particularly toward the end of the lines, may indicate a person is in a state of euphoria or high spirits. The upward line suggests a sense of optimism, energy, and a positive outlook.

Other Indicators include: Large, Flowing Letters; Exaggerated Letter Forms; Consistent Flow.

Graphology does not only focus on slants but also on other traits such as spacing, pressure, size, and shape of letters. These factors together give a more nuanced view of the writer's emotional state. However, it is important to note that while graphology can provide insights into emotional states, it is not a diagnostic tool for clinical conditions like depression or euphoria. These analyses should always be complemented with professional psychological evaluations.

C. Is graphology a pseudoscience?

Graphology is widely considered a pseudoscience. While it involves the analysis of handwriting to infer personality traits or emotional states, there is a lack of scientific evidence supporting its validity or reliability. Here are some key points that contribute to the classification of graphology as a pseudoscience: Lack of Empirical Evidence; No Standardization; Failure to Meet Scientific Criteria; Overgeneralization.

Because of these reasons, graphology is typically regarded as a pseudoscience. While it may offer some insights into a person's emotional or mental state based on patterns in handwriting, its conclusions should be taken with caution and are not a substitute for more established psychological assessments.

In this paper, we will evaluate, from a scientific point of view, whether there is a correlation between depression and the inclination of handwritten text lines. To test this hypothesis, we will use the experimental database Emothaw [15], which consists of several handwritten tasks—including a sentence written in cursive letters—and a DASS evaluation [16].

II. Methods

We estimated the inclination and x-height of all the handwritten sentences and assessed their correlation with the DASS scores, as detailed below.

A. Emothaw database

Emothaw is a publicly available handwriting database designed to link emotional states (depression, anxiety, and stress) with handwriting tasks. The purpose is to aid in detecting negative emotional states through handwriting analysis using machine learning. Key characteristics are:

- 129 subjects aged 21–32 years (mean age: 24.8).
- Emotional states assessed using the DASS-42 questionnaire (Depression Anxiety Stress Scales).
- Participants completed 7 handwriting tasks recorded using a WACOM digitizing tablet with an INTUOS Inkpen: Drawing pentagons, House drawing, Writing four Italian words in cursive capital letters, left-hand loops, right-hand loops, Clock drawing, Copying a sentence in cursive.

Data collected includes pen position, timestamps, pressure, and pen movements both on-paper and in-air.

B. DAS scores: BDI-II and DASS

After completing the handwriting tasks, participants were asked to complete two self-reported questionnaires:

- The Beck Depression Inventory (BDI-II) [17];
- The Depression Anxiety Stress Scales Short Version (DASS-21) [16].

The BDI-II is composed by 21 questions, each one with four answers, being scored on a scale value of 0 to 3. The BDI-II has been developed to measure the severity of depressive symptoms, by considering the last two weeks. When the test is scored, the total score is compared to a key to determine the depression's severity. Four severity levels are defined from minimal range (none) to mild, moderate and severe. This BDI-II questionnaire is based on the fourth edition of the "Diagnostic and Statistical Manual of Mental Disorders" published by the American Psychiatric Association (DSM-IV) and follows its diagnostic criteria for depression.

The DASS-21 is also a self-assessment questionnaire estimating the level of depression, anxiety, and stress over the last week. Each scale consists of seven items, for a total of 21. Participants respond on a 4-point Likert scale (from 0: Did not apply to me at all; to 4: Applied to me very much or most of the time). Each subscale is associated with five severity labels: normal, mild, moderate, severe, and extremely severe. Although widely used in research, the DASS-21 is a screening tool and does not substitute for clinical diagnosis.

C. Handwritting successful experiences for DAS analysis

Several studies [18]–[21] examined differences in terms of handwriting speed between healthy and depressed patients. The results of these studies indicated that depressed patients were significantly slower than healthy ones in both starting the first drawing movement and completing tasks.

The work in [22] broadened the investigation by including, apart from the speed, also features related to space and pressure. The results showed that the depressed patients applied significantly less pressure compared to the healthy ones. In addition, this study confirms that depressed patients were significantly slower than healthy ones. Likewise, the study [23] compared kinematic measures in healthy controls and elderly participants diagnosed with mild major depressive disorder through the clock drawing test. Compared to healthy participants, the clinical group reported significantly lower stroke heights, widths, and lengths, while the number of strokes and temporal measures did not differ between the groups.

The study in [24] shows that it is possible to discriminate between depressed clinical patients and non-clinical participants, exploiting online handwriting and drawing features, obtained from a specific set of tasks. The same methodology was applied in [25] where the investigation was broadened by recruiting a brand-new sample including also participants diagnosed of clinical depression by a mental health-care expert. Results showed that several handwriting/drawing features, can be used to discriminate between depressed and nondepressed subjects.

D. Feature extraction

We have estimated, in addition to the axial pen pressure, the following parameters:

1) Inclination estimation: See [26].
2) x-height: The term x-height refers to the height of lowercase letters, excluding ascenders (like in ”b” or ”d”) and descenders (like in ”p” or ”y”), typically measured from the baseline to the top of the lowercase ”x”. A rotation matrix is applied to remove the skew, aligning the text along the x-axis orientation. This step eliminates any inclination-related distortions in the text. The crests in handwriting refer to the upper extensions of the letters, which rise above the main body (e.g., the upper strokes in letters like ”h” or ”t”). These extensions stretch from the top of the main body to the highest point of the letter. In contrast, the jambs or feet represent the lower extensions, extending below the baseline and filling the space between the bottom of the main body and the lowest point of the stroke (e.g., the descender in ”p” or ”y”) [27]. Figure 1 shows the original text at the top, with a straight line indicating its inclination and center of mass (o). At the bottom, we observe the upper and lower lines of the body text, which result from the following steps:
   - **Centering**: The text is shifted to the origin or coordinates (0, 0).
   - **De-rotation**: A rotation matrix is applied to remove the inclination of the text.
   - **Estimation of Upper and Lower Lines**: Let the online handwritten signal be represented by a set of coordinates (x_i, y_i), where $i = 1, 2, \ldots, n$, and x_i and y_i are the spatial coordinates of the handwriting at each point.
     a) Filtering Data Based on y-Coordinates: The first step is to filter out points that fall outside a specific range defined by the standard deviation of the y-coordinates, in order to exclude the upper (crests) and lower (jambs) extensions of the letters. The filtering condition is:

$$filteredIndices_i = (y_i \geq \mu_y - \sigma_y) \wedge (y_i \leq \mu_y + \sigma_y) \tag{1}$$

     where:

$$\mu_y = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad \sigma_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_y)^2} \tag{2}$$

     The filtered subset of vertical coordinates is then defined as:

$$y_{\text{filtered}} = \{y_i \,|\, \text{filteredIndices}_i = 1\}\,. \tag{3}$$

     b) Estimation of Upper and Lower Baselines After filtering, the upper and lower baselines are estimated as the maximum and minimum values of the remaining y-coordinates:

$$\text{upperLine} = \max\left(\{y_i \mid \text{filteredIndices}_i = 1\}\right), \quad \text{lowerLine} = \min\left(\{y_i \mid \text{filteredIndices}_i = 1\}\right) \tag{4}$$

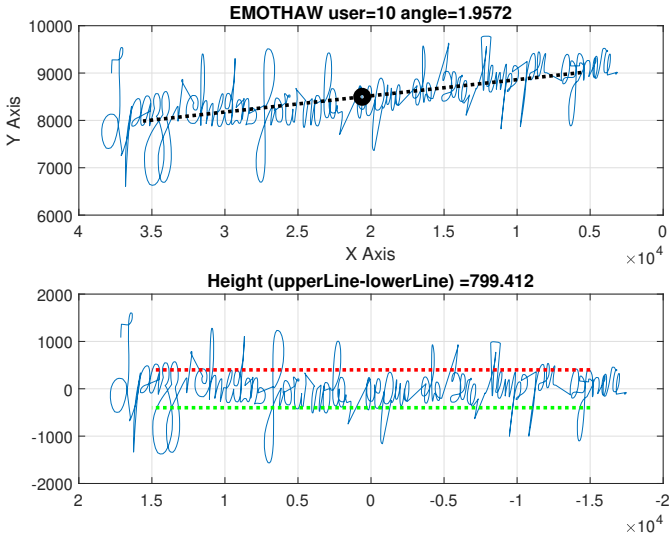


Fig. 1. Original handwritten sentence with a line indicating its inclination (top), center of mass (marked as 'o' at the midpoint of the inclination line), and the result after centering and de-rotation, showing the upper and lower boundaries of the body text (bottom).

III. Results

In our experiments, we used task 7 (handwritten sentence) from the Emothaw database [15]. For these 129 users, we computed the inclination (θ), pressure and body height, which is defined as the difference between the lower and upper boundaries of the body text.

The correlation analysis between handwriting parameters and DASS (Depression, Anxiety, and Stress Scale) scores was performed using the `corrcoef` function in MATLAB, which computes the Pearson correlation coefficient matrix. Given a set of handwriting features (e.g., inclination) and psychological scores (e.g., depression, anxiety, stress), this method quantifies the linear relationship between each feature and each score. Normality was assessed using the Shapiro-Wilk test. When assumptions were violated, Spearman's rho was used instead.

The correlation coefficient, denoted as r, measures the strength and direction of the linear relationship between two variables. It ranges from -1 (perfect negative relationship) to +1 (perfect positive relationship). The strength of the correlation can be interpreted using Cohen's guidelines for effect sizes [28]: small effect ($r \approx 0.10$), medium effect ($r \approx 0.30$), and large effect ($r \approx 0.50$).

In practice, correlation results should be considered in context. For instance, a weak correlation might still be valuable in large datasets, and even a small correlation can have important theoretical implications, particularly when combined with other research findings.

A confidence interval (CI) provides a range of values within which we expect the true population parameter (in this case, the correlation coefficient) to lie, with a certain level of confidence. In this study, we used bootstrapping to estimate a 95% confidence interval for the correlation between two variables. Bootstrapping is a resampling technique that generates multiple samples from the data with replacement, and calculates the statistic (correlation) for each sample. By repeating this process, we obtain a distribution of correlation coefficients. 1000 bootstrap samples were generated using subject-level data vectors (slant, size, etc.), resampling with replacement. Bootstrapping was preferred over t-tests due to potential non-normality and small sample effects in derived metrics.

The 95% confidence interval is derived from the 2.5th and 97.5th percentiles of this distribution. This means that, based on the sample data, we are 95% confident that the true correlation lies within the calculated range. If the interval includes 0, it suggests that the correlation could be weak or nonexistent, indicating uncertainty about the strength of the relationship between the variables. Conversely, if the confidence interval does not include 0, we have stronger evidence of a significant relationship.

In this study, the bootstrapped confidence interval serves to provide a more robust understanding of the variability and uncertainty associated with the observed correlation, complementing the p-value and traditional correlation coefficient. Table I shows the experimental results.

IV. Conclusion

This study scientifically evaluated the correlation between handwriting features—specifically inclination angle and x-height—and emotional states using data from the EMOTHAW database. Our findings reveal no significant correlations between handwriting inclination and emotional states. However, two results are significant with confidence intervals that do not include zero: the correlations between x-height and anxiety, and between mean pressure and anxiety. These results support the broader

TABLE I
CORRELATIONS BETWEEN HANDWRITING FEATURES AND DASS SCORES (DEPRESSION, ANXIETY, AND STRESS)

Correlation	Correlation Value	Confidence Interval	p-value
Inclination vs Depression (D)	-0.194	[-0.298, 0.163]	0.066
Inclination vs Anxiety (A)	-0.090	[-0.318, 0.106]	0.395
Inclination vs Stress (S)	-0.054	[-0.282, 0.107]	0.612
Height vs Depression (D)	0.043	[-0.177, 0.273]	0.688
Height vs Anxiety (A)	**0.216**	**[0.002, 0.396]**	**0.040**
Height vs Stress (S)	-0.023	[-0.210, 0.180]	0.832
Mean Pressure vs Depression (D)	0.028	[-0.244, 0.225]	0.790
Mean Pressure vs Anxiety (A)	**-0.211**	**[-0.396, -0.005]**	**0.044**
Mean Pressure vs Stress (S)	-0.028	[-0.224, 0.135]	0.791

perspective that graphology lacks scientific reliability for psychological assessment, while also highlighting the potential for certain physical features to correlate with emotional states.

ACKNOWLEDGEMENT This work has been supported by Grant PID2023-146644OB-I00 funded by MICIU/AEI/10.13039/501100011033 and FEDER EU.

REFERENCES

[1] S. Kedar, V. Nair, and S. Kulkarni, "Personality Identification through Handwriting Analysis: A Review," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 5, no. 1, 2015.
[2] R. Senatore and A. Marcelli, "A Neural Scheme for Procedural Motor Learning of Handwriting," in *International Conference on Frontiers in Handwriting Recognition*, 2012, pp. 659–664.
[3] M. Pulver, L. Herbert, I. Stirling, and M. Stirling, *The Symbolism of Handwriting*. British Academy of Graphology, 1994.
[4] S. Prasad, V. K. Singh, and A. Sapre, "Handwriting analysis based on segmentation method for prediction of human personality using support vector machine," *International Journal of Computer Applications*, vol. 8, no. 12, pp. 25–29, 2010.
[5] R. Coll, A. Fornés, and J. Lladós, "Graphological analysis of handwritten text documents for human resources recruitment," in *2009 10th International Conference on Document Analysis and Recognition*. IEEE, 2009, pp. 1081–1085.
[6] M. W. Lam, X. Liu, H. M. Meng, and K. K. Tsoi, "Drawing-based automatic dementia screening using gaussian process markov chains," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, January 2018.
[7] J. Mucha, J. Mekyska, Z. Galaz, M. Faundez-Zanuy, K. Lopez-de Ipina, V. Zvoncak, and I. Rektorova, "Identification and monitoring of parkinson's disease dysgraphia based on fractional-order derivatives of online handwriting," *Applied Sciences*, vol. 8, no. 12, p. 2566, 2018.
[8] P. Werner, S. Rosenblum, G. Bar-On, J. Heinik, and A. Korczyn, "Handwriting process variables discriminating mild alzheimer's disease and mild cognitive impairment," *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, vol. 61, no. 4, pp. P228–P236, 2006.
[9] B. Gawda, "Little evidence for the graphical markers of depression," *Percept. Mot. Skills*, vol. 117, no. 1, 2013. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/24422358/
[10] M. Giannini, P. Pellegrini, A. Gori, and Y. Loscalzo, "Is graphology useful in assessing major depression?" *Psychol. Rep.*, vol. 122, no. 2, pp. 398–410, 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29460687/
[11] A. Vels, *Escritura y Personalidad: Las bases científicas de la Grafología aplicadas a la Seleccíon de Personal, a la Psicología clínica, a la Pedagogía, a los Estudios biográficos y a la Investigacíon de la conducta en general*, 9th ed. Barcelona: Herder Editorial, 2010, Translation of title: Handwriting and Personality: The Scientific Foundations of Graphology Applied to Personnel Selection, Clinical Psychology, Pedagogy, Biographical Studies, and Behavioral Research in General. Paperback, 472 p.
[12] A. Mishra, "Forensic graphology: Assessment of personality," *Forensic Research & Criminology International Journal*, vol. 4, no. 1, pp. 9–12, 2017.
[13] A. Poizner, *Clinical Graphology: An Interpretive Manual for Mental Health Practitioners*. Charles C Thomas Pub Ltd, 2012.
[14] R. Cohen, "In france, it's how you cross the t's," *The New York Times*, October 1993. [Online]. Available: https://www.nytimes.com/1993/10/19/business/in-france-it-s-how-you-cross-the-t-s.html
[15] L. Likforman-Sulem, A. Esposito, M. Faundez-Zanuy, S. Clémençon, and G. Cordasco, "Emothaw: A novel database for emotional state recognition from handwriting and drawing," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 273–284, 2017.
[16] S. H. Lovibond and P. F. Lovibond, "Depression anxiety stress scales (dass–21, dass–42) [database record]," 1995. [Online]. Available: https://doi.org/10.1037/t01004-000
[17] A. Beck, R. A. Steer, R. Ball, and R. W. F., "Comparison of Beck depression inventori.es-IA and-II in psychiatric outpatients," *J. Pers. Assess.*, vol. 67, no. 3, pp. 588–597, 2023.
[18] J. J. M. van Hoof, W. Hulstijn, H. van Mier, and M. Pagen, "Figure drawing and psychomotor retardation: preliminary report," *J. Affect. Disord.*, vol. 29, no. 4, pp. 263–266, 1993.
[19] B. Sabbe, W. Hulstijn, J. V. Hoof, and F. Zitman, "Fine motor retardation and depression," *J. Psychiatr. Res.*, vol. 30, no. 4, pp. 295–306, 1996.
[20] B. Sabbe, W. Hulstijn, J. van Hoof, H. G. Tuynman-Qua, and F. Zitman, "Retardation in depression: Assessment by means of simple motor tasks," *J. Affect. Disord.*, vol. 55, no. 1, pp. 39–44, 1999.
[21] R. Mergl, G. Juckel, J. Rihl, V. Henkel, M. Karner, P. Tigges, A. Schroter, and U. Hegerl, "Kinematical analysis of handwriting movements in depressed patients," *Acta Psychiatr. Scand.*, vol. 109, no. 5, pp. 383–391, 2004.
[22] S. Rosenblum, P. Werner, T. Dekel, I. Gurevitz, and J. Heinik, "Handwriting process variables among elderly people with mild major depressive disorder: A preliminary study," *Aging Clin. Exp. Res.*, vol. 22, no. 2, pp. 141–147, 2004.
[23] J. Heinik, P. Werner, T. Dekel, I. Gurevitz, and S. Rosenblum, "Computerized kinematic analysis of the clock drawing task in elderly people with mild major depressive disorder: An exploratory study," *Int. Psychogeriatr.*, vol. 22, no. 3, pp. 479–488, 2010.
[24] G. Raimo, M. Buonanno, M. Conson, G. Cordasco, M. Faundez-Zanuy, G. M. S. Marrone, F. Marulli, A. Vinciarelli, and A. Esposito, "Handwriting and Drawing for Depression Detection: A Preliminary Study," in *Proceedings of 2nd International Conference on Applied Intelligence and Informatics, AII 2022*, 2022, pp. 320–332.
[25] C. Greco, G. Raimo, T. Amorese, M. Cuciniello, G. Mcconvey, G. C. M. Faundez-Zanuy, A. Vinciarelli, Z. Callejas-Carrion, and A. Esposito, "Discriminative Power of Handwriting and Drawing Features in Depression," *International Journal of Neural Systems*, 2023.
[26] M. Faundez-Zanuy, "Analysis of inclinations in genuine signatures and skilled forgeries," in *22nd Conference of the International Graphonomics Society (IGS 2025)*, Polytechnique Montréal, Canada, Jun. 16-19 2025, www.igs2025.org.
[27] J. L. Vásquez, A. G. Ravelo-García, J. B. Alonso, M. K. Dutta, and C. M. Travieso, "Writer Identification Approach by Holistic Graphometric Features using Off-line Handwritten Words," *Neural Computing and Applications*, vol. 32, no. 2, pp. 15733–15746, 2020.
[28] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Routledge, 1988.

Early Detection and Classification of Parkinsonism: Leveraging the Lognormal Model to Aid Clinical Assessment

Karina Lebel[*1,7], Cristina Carmona-Duarte[2], Pierre Blanchet[3,4], Vanessa Bachir[5], Guillaume Seguin de Broin[6], Réjean Plamondon[6]

[1]Faculty of Engineering, Université de Sherbrooke, Sherbrooke, QC, Canada
[2]Instituto Universitario para el Desarrollo Tecnológico y la Innovación, Universidad de Las Palmas de Gran Canaria, Spain.
[3] Faculty of Dental Medicine, Université de Montréal, Montreal, QC, Canada
[4] Neurology Division, Université de Montréal Hospital Centre (CHU Montreal), Montreal, QC, Canada
[5]École d'Optométrie, Université de Montréal, Montreal, QC, Canada
[6]École Polytechnique, Université de Montréal, Montreal, QC, Canada
[7] Research Center on Aging, CIUSSS de l'Estrie - CHUS, Sherbrooke, QC, Canada

[*]karina.lebel@usherbrooke.ca

***Abstract*— Individuals living with Parkinsonism experience motor and non-motor symptoms that impair their daily functioning and quality of life. Early and precise diagnosis is crucial for optimizing care. This study investigates the use of movement attributes to differentiate asymptomatic elderly individuals from those living with Parkinson's disease and explores the potential characteristic differences between two forms of Parkinsonism: idiopathic and atypical. By analyzing rapid scripted strokes, the study demonstrates that lognormal movement attributes reveal differences in timing parameters between asymptomatic elderly and those with parkinsonism. Additionally, it identifies a potential discriminating factor between idiopathic and atypical parkinsonism. These findings p ave t he w ay t o f urther i nvestigation i nto the capabilities of this technic to support clinical diagnosis.**

Keywords*— *Movement Lognormality Signature, Parkinsonism, clinical diagnosis, Lognormal model, motor control.

I. Introduction

Parkinson's disease is a neurodegenerative disorder caused by the degeneration of dopamine-producing neurons in the brain. Individuals living with Parkinson's disease are significantly impacted by the various symptoms in their daily quality of life. Symptoms related to Parkinson's disease are more commonly associated with motor function, including tremor, rigidity, bradykinesia (slowness of movement) and postural instability. However, nonmotor symptoms also affect their quality of life, such as cognitive deficits, mood disorders, and autonomic dysfunction [1]. Although there is currently no cure, symptoms can be managed with medication to improve daily functioning. However, the efficacy of the treatment varies according to a number of factors, among which the specific classification between idiopathic and atypical Parkinsonism is the most important [1] [2]. Idiopathic Parkinsonism is the most common form of Parkinsonism; it generally responds well to traditional dopaminergic treatments like levodopa. On the other hand, atypical Parkinsonism encompasses various disorders with similar symptoms, such as multiple system atrophy and progressive supranuclear palsy, which typically progress more rapidly and are less responsive to traditional Parkinson drugs [3]. Due to these differences, treatment strategies must be promptly adapted to the specific form of Parkinsonism to optimize care and maintain quality of life. Unfortunately, the distinction between these two forms of Parkinsonism pose a challenge as the clinical differences only appear a few years following diagnosis. This delay limits early optimized management, which in turns affects the quality of care. This underscores the urgent need for improved objective assessment to facilitate earlier detection and classification of Parkinson's disease.

Movement signature analysis, herein defined as the analysis of movement attributes during a specific task, has a potential to provide this deeper information not only on the quality of movement performed, but on the quality of the underlying systems involved in movement generation. Among the different approaches listed in the literature for movement signature analysis are the models based on the Kinematics Theory (KT). KT stipulates that a straight rapid movement reflects the lognormal impulse response of a neuromuscular system, which is made up of the combined contributions of numerous interconnected subsystems. The lognormal model was developped by Plamondon in 1995, and later formalized by Plamondon *et al.* in 2003 [4] [5] [6] [7]. This model, based on the Kinematics Theory, mathematically captures the characteristics of a signature to reflect the physiological organization and execution of the movement. The most common application reside in the analysis of a straight scripted trait realized on a digital tablet and analyzed using the Lognormal Model. For years now, analyzing the velocity profiles of rapid scripted movements using this method has provided insights into the underlying mechanisms of movement generation. A variation of this model, the Sigma-Lognormal model, was also adapted to investigate cranio-caudal

signature during a turn [8]. This approach enabled the distinction between individuals with early onset of Parkinson's disease and healthy elderly individuals, as well as to detect the impact of medication on the quality of the turn for patients living with Parkinson's disease [9]. However, the clinical acceptability of turn signature analysis is limited for daily use. There is thus an increasing interest to develop a simple tool to aid clinicians in the early diagnostic of Parkinson's disease, and the classification of the specific type of Parkinsonism in order to better adjust medication regimen. This pilot study explores the potential of analyzing rapid scripted traits using the Lognormal model in order to differentiate individuals with Parkinsonism from asymptomatic elderly, and to further distinguish between idiopathic and atypical Parkinsonism.

II. FRAMEWORK & METHODS

This section provides a concise overview of the assessment system, the lognormal model analysis parameters, and the methodology employed in this study.

A. Scripted Traits System Assessment

The first system used for this study, also refered to as the pen stroke test, is a lognometer consisting of a Wacom tablet pre-programmed with a series of scripted tests . The tablet is connected to a laptop, enabling real-time monitoring of test progression and an initial quality check. A trial is considered valid if the minimum conditions are met, specifically a stable initial pen position and stabilization within the end zone.

The protocol includes tasks such as rapid straight line with visual and auditory cues, a 10-second triangular trajectory, a spiral task, a maximum oscillation task, and a task where the participant writes in a repeated sequence of e-l in a continuous manner. The primary test of interest for this paper is the rapid straight line with visual cues. In this test, the participant initially positions the stylus in one of the designated side zones. When the tablet turns green, the participant must reach the opposite side zone as quickly as possible. This test is repeated 30 times consecutively.

Additional tasks involving upper limb movements, voice assessment, and eye tracking were also performed by participants using various equipment. This multimodal approach aims to evaluate the impact of the disease from multiple perspectives while leveraging the potential for complementary analysis. However, the analysis of these signatures, intended to be conducted using the same lognormal model to support complementarity, is beyond the scope of the present article.

B. Lognormal Model

The lognormal model allows for the objective assessment of movement trajectories, accounting for both normal and pathological variability. Based on Kinematics Theory, this model can be seen as a reverse engineering method to reconstruct movements and extract mathematical parameters linked to the human neuromotor system. Movements are composed of one or more strokes, each characterized by the following lognormal parameters:

- t_0: This timestamps corresponds to the time at which the central nervous system emits the motor command. It represents the central nervous system reaction time following the stimulus, not to be confused with the mechanical reaction time.
- D: The amplitude of the lognormal stroke, corresponding to the total distance traveled due to this specific movement primitive.
- μ: The time required to reach the median of the curve on a logarithmic scale, linked to the actual reaction speed.
- σ: The response time on a logarithmic scale, characterizing the duration of the motion.

These parameters allow for the reconstruction of a lognormal stroke using the equation 1. The sigma-lognormal variation of the model considers the contribution of agonist strokes, which facilitate the execution of movement, and antagonist strokes, providing the resistance to stabilize or stop the motion.

$$\overrightarrow{v}(t) = \Sigma_{i=1}^{n} \overrightarrow{v}_i(t; t_{0i}; \mu_i, \sigma_i^2) = \Sigma_{i=1}^{n} D_i \begin{bmatrix} cos(\theta_i(t)) \\ sin(\theta_i(t)) \end{bmatrix} \Lambda_i(t; t_{0i}; \mu_i, \sigma_i^2) \tag{1}$$

C. Methodology

Participants living with Parkinson's disease were recruited from the CHUM or Réseau Parkinson Quebec. Inclusion criteria included a diagnosis of parkinsonism, either idiopathic or atypical, motor symptoms of parkinsonism for no more than six years, age between 50-75 years, and the ability to leave home safely in the morning without having taken antiparkinsonian medication for the last 12 hours. Exclusion criteria included major neurocognitive disorders, a history of other neurological conditions (e.g., hemorrhagic stroke), a history of oromandibular or laryngeal procedures, uncorrected deafness, and any contraindication to pupillary dilation. The study aim to recruit a total of 60 individuals with parkinsonism (30 with idiopathic parkinsonism, 30 atypical parkinsonism) and 30 age- and gender-matched volunteers.

Data are collected during a single visit to the UdeM School of Optometry. After providing informed consent, participants undergo a clinical evaluation, including section III of the Unified Parkinson Disease Rating Scale (UPDRS) for motor impact severity, and the Radboud Oral Motor Inventory for Parkinson's Disease (ROMP). All participants are examined on their OFF state, meaning approximately 12h after their last antiparkinsonism medication intake. Performance assessment is conducted in

TABLE I. DIFFERENCIATION BETWEEN SCRIPTED STROKES

Parameter	Asymptomatic Strokes *n = 414*	Parkinson Strokes *n = 692*	p
Agonist Parameters			
t_0	322.1 ± 107.4	344.1 ± 111.4	**p = 0.002***
D	265.53 ± 108.38	272.86 ± 83.93	p = 0.208
μ	-1.48 ± 0.31	-1.24 ± 0.41	**p < 0.001***
σ	0.52 ± 0.16	0.53 ± 0.13	p = 0.655
Antagonist Parameters			
t_0	357.3 ± 153.9	408.3 ± 176.7	**p < 0.001***
D	87.47 ± 49.38	85.52 ± 43.75	p = 0.713
μ	-1.20 ± 0.32	-0.95 ± 0.37	**p < 0.001***
σ	0.26 ± 0.11	0.28 ± 0.11	**p = 0.003***

TABLE II. PARKINSONISM VS ASYMPTOMATIC ELDERLY INDIVIDUALS

Parameter	Asymptomatic Elderly *n = 18*	Parkinson *n = 25*	p
Global Parameters			
SNR	20.68 ± 2.93	19.3 ± 2.1	p = 0.077
Nb Strokes	23.0 ± 10.0	27.7 ± 5.5	p = 0.055
Agonist Parameters			
t_0	331.9 ± 57.9	348.6 ± 69.1	p = 0.414
D	297.28 ± 135.22	277.18 ± 63.30	p = 0.526
μ	-1.46 ± 0.33	-1.24 ± 0.38	p = 0.052
σ	0.55 ± 0.14	0.54 ± 0.11	p = 0.759
Antagonist Parameters			
t_0	357.3 ± 153.9	408.3 ± 176.7	**p < 0.001***
D	87.47 ± 49.38	85.52 ± 43.75	p = 0.713
μ	-1.20 ± 0.32	-0.95 ± 0.37	**p < 0.001***
σ	0.26 ± 0.11	0.28 ± 0.11	**p = 0.003***

four distinct phases: (i) scripted tasks, (ii) upperlimb movement tasks, (iii) vocal performance tests, and (iv) occulomotor tests. This paper focuses on the first set of tasks. More specifically, parameters obtained from the rapid straight line scripted task with visual cues are extracted and analyzed.

III. RESULTS

This section reports the results from a preliminary analysis conducted on participants recruited between 2022 and 2024. A total of 45 participants were recruited, including 20 asymptomatic elderly individuals (18 with valid data) and 25 individuals with parkinsonism (4 of whom had atypical parkinsonism diagnosis).

A. Parkinson vs Asymptomatic Detection

The first level of analysis focuses on distinguishing parkinsonism from asymptomatic elderly individuals. This is achieved by examining parameters derived from the analysis of rapid scripted movements using the Sigma-Lognormal model. The initial analysis examines the results at the scripted stroke level to identify apparent differences in the characteristics of a Parkinsonian stroke compared to the asymptomatic elderly. Following initial data collection, a quality control procedure was applied. Pen strokes were excluded if they contained missing timestamps, recording artifacts, or incomplete trajectories unsuitable for lognormal modeling. This quality filtering accounts for the reduced number of tracks included in the final analysis.

Table I presents analyses at the stroke level, where each valid pen stroke is considered independently. Recall that the difference between the number of collected and analyzed strokes is due to the initial quality control procedure, which excluded invalid or incomplete data. All variables are assessed using t-tests for independant groups. The next level of analysis involves defining representative parameters for each person based on the multiple strokes performed and analyzed. Table II presents data aggregated at the participant level by averaging all valid strokes for each individual. This aggregation reflects the clinical objective of identifying individuals across different groups rather than analyzing isolated movements.

B. Idiopathic vs Atypical Parkinsonism

The second objective of this study consists in exploring the differences in motor control between idiopathic and atypical parkinsonism. Given the limited number of participants with atypical parkinsonism, analyses were conducted at the stroke level. Parameters related to agonist and antagonist lognormals were analyzed. Based on the previous section, the parameter t_0, namely the delay between the visual cue and the onset of movement, μ, which provides a measure of central tendancy in the velocity profile, and σ, associated to the time required to perform the movement, were investigated. Figure 1 illustrates the distribution of the parameters, along with the statistical relationships identified using a one-way ANOVA.

IV. DISCUSSION & CONCLUSION

The first objective of this paper was to identify differences in the intrinsic parameters that describe the velocity profile in a scripted rapid stroke test with visual cues to discriminate between asymptomatic elderly individual and people living with Parkinsonism. Analyses performed at stroke level confirmed significant difference in the various timing parameters (t_0, μ, σ), but not in the distance parameters (D). This may be explained by the accessibility of the task performed, with the required distance being within the capability of all participants but performed differently in terms of timing. An interesting observation arises when analyzing the data at the individual level. The number of successful strokes analyzed is almost significantly lower for asymptomatic elderly individuals, suggesting that this group may not be as focused on the task as individuals living with Parkinsonism. In addition, the timing parameters are significantly higher for the antagonist portion of the movement. The trends observed within the agonist parameters are similar to the previous analyses at stroke level but are not yet significant.

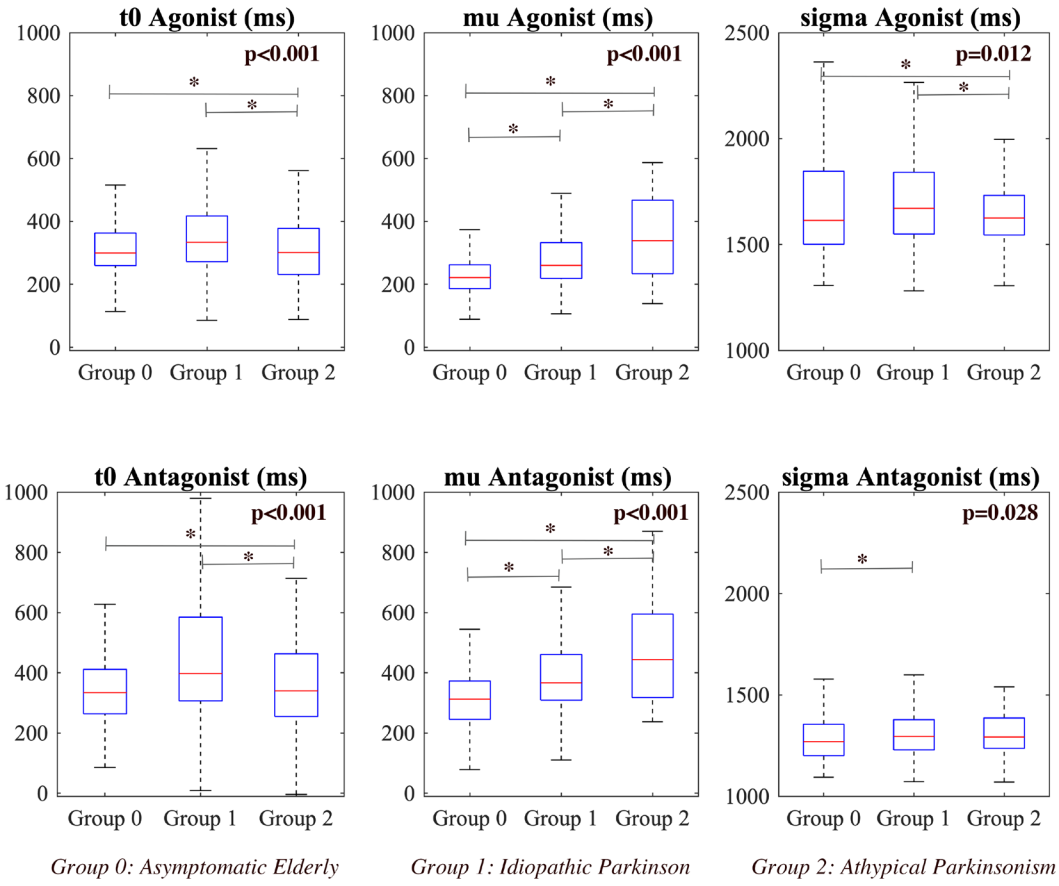


Fig. 1. Exploration of differences between Idiopathic and Aypical Parkinsonism

Further analysis is needed to determine whether the observed difference in results is due to insufficient statistical power or if averaging parameters across all strokes performed by an individual is the most appropriate method for characterizing them. In this regard, future work will examine within-individual stroke variability to assess potential differences in consistency.

An exploratory analysis was conducted to identify emerging trends between idiopathic and atypical Parkinsonism. Due to the limited number of atypical participants, the analysis was performed at the stroke level, and caution is advised in interpreting these results. Nevertheless, the three timing parameters appear statistically discriminant, with t_0 and μ being the most revealing. These results suggest that both forms of Parkinsonism involve slower movements, though the delay in initiation of motion for atypical Parkinsonism could be less than idiopathic Parkinsonism. The time required to reach the median of the motion is a parameter to closely monitor, as preliminary results show an increasing trend from asymptomatic elderly to idiopathic Parkinson, reaching the highest value for atypical Parkinsonism strokes. Variability in stroke repetitions, across individuals and over time, will also be further analyzed to explore potential distinctions in patterns between idiopathic and atypical Parkinsonism.

All participants in this study completed multiple motor assessment tasks, including handwriting, spiral drawing, voice recordings, and eye movement tracking. Future research will leverage this multimodal dataset to explore the combined diagnostic potential of these different evaluation modalities. In particular, upcoming analyses will assess how integrating data from these tasks can enhance our understanding of the sigma-lognormal model's clinical relevance and improve the detection and characterization of Parkinsonian motor symptoms.

Acknowledgment

This manuscript was reviewed for language and fluidity with the assistance of Microsoft Copilot. The authors would like to thank Réseau Parkinson Québec (RPQ) for their help during recruitment. This work was supported by the CIRCA funding "Apport de la théorie cinématique dans le diagnostic différentiel précoce de la maladie de Parkinson", the Spanish project PID2021-122687OA-I00 / AEI / 10.13039 / 501100011033 /FEDER, UE, and CAS22/00270, and the Canada Research Chair in Biomechanical Movement Signature to Better Understand and Reduce Disability.

References

[1] S. A. Schneider and C. Klein, "The Many Faces of Parkinsonism," pp. 117–128, 2013.
[2] J. Ridgeway-Diaz and L. Marsh, "Parkinson's Disease and Atypical Parkinsonian Syndromes," pp. 298–C17.P290, 2022.
[3] E. Schäffer and D. Berg, "Redefinition of Parkinson's Disease," *Neurology International Open*, vol. 1, no. 02, pp. E65–E70, 2017.
[4] R. Plamondon, "A kinematic theory of rapid human movements. Part II. Movement time and control." *Biological cybernetics*, vol. 72, no. 4, pp. 309–320, 1995.
[5] R. Plamondon, "A kinematic theory of rapid human movements. Part I. Movement representation and generation." *Biological cybernetics*, vol. 72, no. 4, pp. 295–307, 1995.
[6] R. Plamondon, "A kinematic theory of rapid human movements: Part III. Kinetic outcomes." *Biological cybernetics*, vol. 78, no. 2, pp. 133–145, 1998.
[7] R. Plamondon, C. Feng, and A. Woch, "A kinematic theory of rapid human movement. Part IV: a formal mathematical proof and new insights." *Biological cybernetics*, vol. 89, no. 2, pp. 126–138, 2003.
[8] K. Lebel, H. Nguyen, C. Duval, R. Plamondon, and P. Boissy, "Capturing the Cranio-Caudal Signature of a Turn with Inertial Measurement Systems: Methods, Parameters Robustness and Reliability." *Frontiers in bioengineering and biotechnology*, vol. 5, p. 51, 2017.
[9] K. Lebel, C. Duval, H. P. Nguyen, R. Plamondon, and P. Boissy, "Cranio-Caudal Kinematic Turn Signature Assessed with Inertial Systems As a Marker of Mobility Deficits in Parkinson's Disease." *Frontiers in neurology*, vol. 9, p. 22, 2018, Jan. 29.

Deep Temporal Modeling for Automated Differentiation of Parkinson's Disease and Progressive Supranuclear Palsy Based on Gait Analysis

Cesare Davide Pace*, Maria Gaglione †, Alessandro Marco De Nunzio‡§, Claudio De Stefano*, Francesco Fontanella*, Maria Francesca De Pandis†x, Mario Molinara*x

*Università degli Studi di Cassino e del Lazio Meridionale,
Via Gaetano di Biasio, 43, 03043 Cassino FR, Italy
{cesaredavide.pace, destefano, fontanella, m.molinara}@unicas.it

†San Raffaele Cassino Rehabilitation Institute, Cassino, Italy
{maria.gaglione, francesca.depandis}@sanraffaele.it

‡Department of Research and Development, LUNEX, Luxembourg
§Luxembourg Health & Sport Sciences Research Institute ASBL, Luxembourg
adenunzio@lunex.lu
xBoth authors contributed equally to this work.

***Abstract*— The classification of neurodegenerative diseases based on movement analysis has gained increasing attention in recent years. In this work, we investigate the application of deep learning-based temporal modeling techniques to distinguish between Parkinson's Disease (PD) and Progressive Supranuclear Palsy (PSP). Unlike traditional approaches that rely on hand-crafted spatiotemporal and kinematic features, our methodology directly processes raw 3D marker trajectories obtained from BTS Gait Systems during walking and standing trials. This end-to-end learning pipeline captures disease-specific motion patterns without explicit feature engineering. While previous studies have explored similar architectures for classifying PD versus healthy controls, their effectiveness in distinguishing PD from PSP using raw motion data remains underexplored. Our results demonstrate that integrating deep temporal models within gait analysis frameworks provides a promising direction for automated disease classification.**

Keywords*— *Gait analysis, Parkinson's Disease, Progressive Supranuclear Palsy, Motion Capture, Temporal Deep Models

I. Introduction

Gait disturbances are among the earliest and most disabling symptoms of movement disorders, such as Parkinson's Disease (PD) and Progressive Supranuclear Palsy (PSP). Distinguishing PSP from PD is of critical importance due to their differing disease progression, therapeutic responses, and prognoses. Unlike PD, PSP progresses more rapidly and exhibits limited response to dopaminergic therapy, necessitating alternative management strategies. Furthermore, early and accurate differentiation enables better patient counselling, targeted interventions, and enrollment in clinical trials for disease-modifying treatments [1]. Traditional diagnostic approaches rely on clinical rating scales and subjective observations, which may result in misclassification, particularly in the early stages of the disease. [2].

Advancements in motion capture technologies, such as BTS Gait Systems, have facilitated objective gait analysis, enabling a more precise differentiation of movement disorders. Recent studies have demonstrated the efficacy of gait analysis in distinguishing PSP phenotypes and differentiating PD from PSP using wearable sensors and machine learning [3], [4]. However, most existing methods extract predefined spatiotemporal and kinematic features from gait data, which may overlook subtle temporal dependencies in movement patterns [5], [6].

In this work, we explore the use of deep temporal models like Multi-Head Self-Attention (MHSA), Long Short-Term Memory (LSTM), and Recurrent Neural Networks (RNN) to classify PD and PSP directly from raw 3D motion marker data. Unlike previous studies focusing on feature extraction, our approach leverages end-to-end learning to detect discriminative motion patterns automatically. This builds upon prior research that applied these temporal architectures for PD classification [7]–[9] and deep time-series models for gait anomaly detection [6].

Our study aims to contribute to integrating automated classification tools within gait analysis systems, providing clinicians with an objective, data-driven diagnostic aid. By avoiding manual feature engineering, our approach captures complex temporal dependencies inherent in gait disturbances, potentially improving diagnostic accuracy. The remainder of this paper presents our Material II, Methods III, Results IV and Conclusion V.

II. Materials

A. Participants

The dataset consists of 13 subjects: 7 with PD and 6 with PSP. The gender distribution within each group is: PD – 4 females, 3 males; PSP – 5 females, 1 males. The subjects were recruited from San Raffaele Cassino institute, where diagnoses were made by experienced clinicians following established diagnostic criteria, and confirmed through clinical evaluations.

B. Data Acquisition

Gait data were collected in a controlled laboratory environment using the BTS Gait System, a high-precision optical motion capture system. Each subject was subjected to multiple standing and walking tasks, following the Davis protocol, which ensures standardized marker placement and motion tracking procedures. A total of 79 trials were recorded across all subjects, which were subsequently used to train the deep learning architecture. Data collection followed the Davis marker set, with markers placed at anatomical landmarks, including the sacrum, pelvis, thighs, knees, ankles, feet, shoulders, and upper spine. The sampling frequency of the BTS Gait System was set to 120 Hz, ensuring high temporal resolution for motion analysis.

C. Preprocessing

Before analysis, raw data were cleaned to remove missing values (NaN) and zero values, which usually occur when a subject moves outside the acquisition zone. This step ensured that only valid motion sequences were used in the modelling process. Since gait sequences often vary in length, a fixed-length window of 3000 samples was extracted from each sequence. This window was centred on the middle of the sequence, assuming that this region contains the most informative content regarding gait dynamics. By standardizing the input length, this approach ensures consistency in the model input representation, reducing the variability caused by differences in walking speed and trial duration.

III. Methods

A. Model Architecture

Our proposed model follows an end-to-end deep learning approach to process raw 3D gait motion data, and an architecture scheme is shown in Fig. 1.

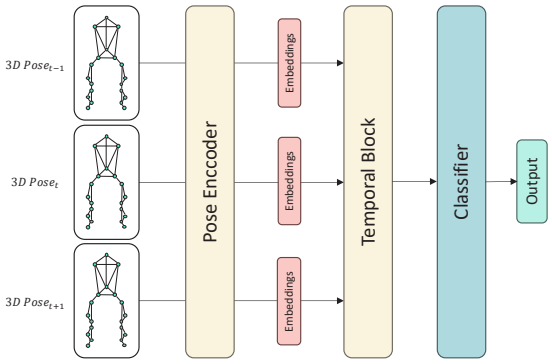


Fig. 1. Scheme of the proposed approach.

The architecture consists of three main components:

- **Pose Encoder:** The input, consisting of raw 3D marker data, is first transformed into a latent space using a fully connected network. This step ensures that relevant spatial features are extracted before temporal modelling.
- **Temporal Block:** To capture the temporal dependencies in gait dynamics, we employ recurrent or attention-based sequence modelling. This component is responsible for modelling long-range dependencies and distinguishing subtle movement patterns characteristic of PD and PSP. Recurrent Neural Network(RNN) [10], Long Short-Term Memory(LSTM) [11], Multi-Head Self-Attention(MHSA) [12] with Positional Encoding(PE) are used as Temporal Block.
- **Classification Head:** The final output of the temporal block is passed through a fully connected network that maps features from the temporal block to classification probabilities.

B. Evalutation metrics and Cross-Validation Strategy

To evaluate the model's ability to differentiate PD from PSP, we use the accuracy, sensitivity, specificity, F1-Score and Matthews correlation coefficient (MCC) metrics. Given the limited sample size, we employed a Leave-One-Subject-Out Cross-Validation strategy to ensure robust performance estimation. In this approach, each subject was iteratively left out as a test case while the model was trained on the remaining subjects. Within each training fold, the remaining data was further split into 80% for training and 20% for validation, ensuring a structured approach to model optimization and performance tuning.

IV. Results

The preliminary results, reported in Tab. I indicate that the MHSA with the Positional Encoding approach achieved the highest classification performance, with an accuracy of 0.769, an F1-score of 0.723, and the best MCC value of 0.534. This suggests that using MHSA mechanisms enhances the ability to capture relevant temporal dependencies in gait sequences, making it the most effective method for the classification of PD and PSP.

The RNN-based approach performed moderately well, achieving an accuracy of 0.679 and an MCC of 0.355, but its sensitivity was particularly low (0.193), indicating that it struggles with capturing key distinguishing features in certain cases.

On the other hand, the LSTM model underperformed significantly, with an accuracy of only 0.4103, a negative MCC value (-0.396), and a sensitivity of 0.0, meaning it failed to correctly identify PSP cases. This suggests that LSTM alone might not be well-suited for this specific classification task, possibly due to limitations in handling the complexity of raw 3D motion marker data without additional attention mechanisms.

TABLE I. Performance comparison of different temporal models for PD vs PSP classification.

Methods	Accuracy	Sensitivity	Specificity	F1-Score	MCC
Recurrent Neural Network (RNN)	0.679	0.194	**1.000**	0.557	0.355
Long Short-Term Memory (LSTM)	0.410	0.000	0.681	0.291	-0.396
Multi-Head Self-Attention (MHSA) + Position Encoding	**0.769**	**0.452**	0.979	**0.723**	**0.534**

V. Conclusion

The preliminary results highlight the effectiveness of different temporal models in distinguishing between PD and PSP based on gait analysis. MHSA with Positional Encoding outperformed other methods, achieving the highest accuracy and F1-Score, demonstrating its ability to capture complex temporal dependencies in motion data. The RNN performed moderately but exhibited lower sensitivity, indicating limitations in capturing subtle gait variations. Meanwhile, LSTM struggled significantly, suggesting that LSTM alone may not be sufficient for this classification task. These findings suggest that attention-based architectures, particularly MHSA, offer a promising approach to improving automated gait-based classification of neurodegenerative diseases. While the model shows promising results, the small dataset confounders may affect generalizability. Future work will explore a larger dataset to assess model performance and more sophisticated temporal or recurrent models.

References

[1] R. Liscic, K. Srulijes, A. Groeger, W. Maetzler, and D. Berg, "Differentiation of progressive supranuclear palsy: clinical, imaging and laboratory tools," *Acta Neurologica Scandinavica*, vol. 127, no. 5, pp. 362–370, 2013.

[2] M. Picillo, C. Ricciardi, M. F. Tepedino, F. Abate, S. Cuoco, I. Carotenuto, R. Erro, G. Ricciardelli, M. Russo, M. Cesarelli *et al.*, "Gait analysis in progressive supranuclear palsy phenotypes," *Frontiers in Neurology*, vol. 12, p. 674495, 2021.

[3] N. Pisani, C. Ricciardi, M. Picillo, F. Abate, A. R. Avallone, F. Amato, and M. Cesarelli, "Using wearable sensors and motion parameters for recognizing progressive supranuclear palsy phenotypes," in *2023 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*. IEEE, 2023, pp. 864–869.

[4] M. De Vos, J. Prince, T. Buchanan, J. J. FitzGerald, and C. A. Antoniades, "Discriminating progressive supranuclear palsy from parkinson's disease using wearable technology and machine learning," *Gait & Posture*, vol. 77, pp. 257–263, 2020.

[5] C. Ricciardi, M. Amboni, C. De Santis, G. Improta, G. Volpe, L. Iuppariello, G. Ricciardelli, G. D'Addio, C. Vitale, P. Barone *et al.*, "Using gait analysis' parameters to classify parkinsonism: A data mining approach," *Computer methods and programs in biomedicine*, vol. 180, p. 105033, 2019.

[6] G. Paragliola and A. Coronato, "Gait anomaly detection of subjects with parkinson's disease using a deep time series-based approach," *Ieee access*, vol. 6, pp. 73 280–73 292, 2018.

[7] A. H. Butt, F. Cavallo, C. Maremmani, and E. Rovini, "Biomechanical parameters assessment for the classification of parkinson disease using bidirectional long short-term memory," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 5761–5764.

[8] F.-C. Wang, Y.-C. Li, T.-Y. Kuo, S.-F. Chen, and C.-H. Lin, "Real-time detection of gait events by recurrent neural networks," *IEEE Access*, vol. 9, pp. 134 849–134 857, 2021.

[9] H. X. Tan, N. N. Aung, J. Tian, M. C. H. Chua, and Y. O. Yang, "Time series classification using a modified lstm approach from accelerometer-based data: A comparative study for gait cycle detection," *Gait & posture*, vol. 74, pp. 128–134, 2019.

[10] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[11] S. Hochreiter, "Long short-term memory," *Neural Computation MIT-Press*, 1997.

[12] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

Modeling Action Potentials with Sigma-Lognormal Equations

Réjean Plamondon[1], Moussa Djioua[1] and Christian O'Reilly[2-5]

[1]*Génie Électrique, Polytechnique Montréal, 2500 Chemin de Polytechnique, Montréal, H3T-1J4, Québec, Canada*
m.djioua@polymtl.ca , rejean.plamondon@polymtl.ca

[2] Computer Science and Engineering, University of South Carolina, Columbia, SC, USA
christian.oreilly@sc.edu

[3] Artificial Intelligence Institute, University of South Carolina, Columbia, SC, USA

[4] Carolina Autism and Neurodevelopment Research Center, University of South Carolina, Columbia, SC, USA

[5] Institute for Mind and Brain, University of South Carolina, Columbia, SC, USA

***Abstract*—** In the context of investigating the ubiquity of lognormal impulse responses in humans and animals, this paper frames the voltage clamp results obtained by Hodgkin and Huxley (HH) in their squid giant axon experiments in a new light using the Kinematic Theory and its lognormality convergence. The stereotypical waveform of the action potential (AP), mainly built by sodium and potassium conductance changes, is modeled by a Sigma-Lognormal function without needing solving differential equation systems. This approach models the AP with three lognormal equations, representing the sodium conductance, the potassium conductance pattern, and ionic leak-age currents in voltage space. This parametric modeling approach is supported by best curve-fitting results of intracellular action poten-tial data recorded *in vitro,* and by models and observations reported in multiple neurophysiological studies. Through the four lognormal parameters, one can analyze and/or synthesize the conductance behavior of each current. This alternative formulation, derived from modeling basic biological principles, can serve as a baseline computational tool for neuroscience applications, such as implementing spik-ing neural network models and spike sorting.

Keywords—Action Potential, Neuroscience, Kinematic Theory, Lognormality, Sigma-Lognormal model

I. Introduction

Lognormal distributions are applied to fit statistical and probabilistic data in numerous applications in economics, business, physics, chemistry, industry, manufacturing, biology, ecology, atmospheric sciences, geology, and geography [1,2,3]. In neurosciences, Buzaki and Mizuseli [4] have made a survey reporting many physiological and anatomical lognormal parameter distributions observed in brain studies. These statistical models rely on the central limit theorem predicting the emergence of lognormal distributions when products of many independent random variables or sums of many proportionally coupled random variables describe stochastic processes. The use of lognormal functions as deterministic and causal impulse responses of dynamical systems rather than statistical distributions of random variables is less known but of growing interest. Indeed, the concept of lognormal impulse response has been considered to model human movements with many coupled neuromuscular subsystems under a time delay proportionality constraint using the central limit theorem [5,6,7,8]. Since then, lognormal impulse responses have been used to characterize fingers, wrist, hand, arm, eyes, head, and trunk movements, as well as speech, gait and electrocardiogram (ECG) without forgetting applications in fish behaviors, pets, and livestock, biomarkers for ADHD, concussions, muscular fatigue, Parkinson and Alzheimer diseases to name a few [9,10,11].

According to this paradigm, any dynamical system that can be modelled by convolving a large number (in theory, an infinite number, but in practice a number of 30 is often sufficient) of proportionally coupled subsystems will have a lognormal impulse response, which suggests that lognormality might reflect auto similarity in nature. To foster this hypothesis regarding the ubiquity of lognormal phenomena, we investigate in this paper the corner stone of all these neuromotor phenomena: the action potential (AP). In other words, this paper explores a brand-new field of applications regarding the Kinematic Theory and its lognormality principle: can it be used to model the action potential waveform? This question that was first raised years ago in [12].

The action potential (AP) constitutes the stereotypical response of neurons and other excitable cells depolarized above their firing threshold. It is a fundamental mechanism for neurons to communicate between themselves, control the behavior of other cells (e.g., trigger the contraction of muscular fibers), and encode physical quantities perceived by sensory cells. The seminal work of Hodgkin and Huxley has provided the basis for understanding the AP genesis [13,14,15,16,17]. They showed that the AP results from the interactions of dynamic voltage-dependent ionic currents causing avalanche reactions depolarizing and repolarizing the cell mem-brane in quick succession. This process is initiated when the membrane potential is depolarized enough to trigger the opening of sodium (Na^+) ion channels, causing a large influx of positive charges, further depolarizing the membrane, and triggering more Na^+ channels to open. As the membrane potential rises to its apex, Na^+ channels close due to their inactivation at high potential. Simulta-neously, voltage-sensitive potassium (K^+) opens in large quantities, causing a large efflux of positive charges and bringing the mem-brane potential back to its resting potential, often overshooting and causing an afterhyperpolarization. In first approximation, we can approximate this process by modeling it as a ballistic open-loop process caused by a triggering event (i.e., the membrane potential reaching the response threshold). This description of the AP generation process shows a direct relationship between the AP waveform and the conductance of the membrane for the Na^+ and K^+ ions. The HH formalism describes the AP as follows:

$$C_m \frac{dV}{dt} = -(V - V_k)g_k - (V - V_{Na})g_{Na} - (V - V_l)g_l \quad (1)$$

where C_m stands for the membrane capacity, V for the membrane potential from its resting value, V_k, V_{Na} , and V_l for the equilibrium potentials, and g_k, g_{Na}, and g_l for voltage-dependent conductance changes for the K^+, Na^+, and leak currents, respectively. The conductance change is considered to result from the probability of ion channel opening, with a voltage-dependent time constant at the population dynamics level. It was modeled as a power law of functions expressed as a differential equation. However, as pointed out by the authors, this equation is not the unique solution to modeling AP, and other mathematical representations can fit the experimental data [18,19, 20,21].

In the context of this diversity of AP models, we propose the Sigma-Lognormal model as a simple, analytical and operational representation of the AP profile instantiated when the membrane potential crosses the response threshold and modeled as a discrete event generally represented using a Dirac function. As will be seen, this new waveform modeling is supported by an excellent fit, measured by a signal-to-noise ratio (SNR) and expressed in dB, for intracellular AP recorded *in vitro* on Lamprey neurons, and by physiological and mathematical arguments reported in different studies dealing with the HH model. The lognormal parameters extracted from the reconstruction of an AP provide a new window to analyze this neural waveform. Among other things, information regarding the time occurrence, the logtime delays, and the logresponse times of the ionic currents involved in the AP process can be quantified and studied.

II. Modeling considerations

A. Introducing the Sigma-Lognormal model of AP

Describing the AP using the Kinematic Theory, widely used in human movement science [5,6,8,9], is motivated by the assumption that the AP arises from transmembrane movements of Na^+ and K^+ ions, causing changes in ionic concentrations with asymmetric bell-shaped profiles. Initially, a shallow depolarization triggers a Na+ influx, which results in a deeper depolarization and, in turn, a K+ efflux. The AP can be seen as resulting from these ionic kinetics.

The modeling of time series as sets of overlapping lognormal components is rooted in the Kinematic Theory and its lognormality principle [9]. The cornerstone of this theory is that the asymmetric bell-shaped profile observed in the causal impulse responses of biological systems is well-described by a lognormal function. When impulse responses involve synergies between two or more subsystems, we can describe their profiles by a vectorial summation of lognormal equations known as the Sigma-Lognormal model. We find a similar synergy between the Na^+ and K^+ currents for the AP. However, transmembrane leakage currents from different sources are also present in neurons, and we use a third lognormal component to capture this contribution.
The lognormal function, which describes the causal impulse response of a system coupled proportionally, is formally given by:

$$\Lambda(t; t_0, \mu, \sigma) = \frac{1}{\sigma(t-t_0)\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}[\log(t-t_0)-\mu]^2} \quad (2)$$

where t_0 is the time occurrence of a discrete event, μ is a time delay, and σ is a time response, expressed on a log-time scale. Specially, the voltage response $V(t)$ made up of the contribution of multiple ionic currents is defined by the following equation:

$$\mathrm{V(t)} = \sum_{j=1}^{3} \mathrm{A_j} \Lambda\left(t; t_{0_j}, \mu_j, \sigma_j\right) \quad (3)$$

where A_j is a current density amplitude and the polarity of the current is coded in the sign of A_j .

B. Mathematical considerations

HH equations successfully describe conductance changes in transmembrane Na^+ and K^+ ion channels during an AP. However, many different models [24,25] have been proposed to improve the description of the experimental data with simple mathematical expressions. Particularly, [25] has suggested that the following second-order differential equation might fit better the experimental data than the first-order equations used in HH equations.

$$\frac{d^2V_j}{dt^2} + \gamma_j \frac{dV_j}{dt} + \delta_j V_j = 0 \quad (4)$$

where the subscript j stands for the Na^+ or K^+ ions, and γ_j, δ_j are coefficients. Interestingly, the lognormal equation (2) obeys a second-order differential equation similar to (4) and provides an analytical interpretation of its parameters. Indeed, by using $V_j = D_j \Lambda\left(t; t_{0_j}, \mu_j, \sigma_j\right)$ in equation (4), one can find:

$$\gamma_j = \frac{1}{t-t_{0_j}}\left[1 + \frac{\log\left(t-t_{0_j}\right)-\mu_j}{2\sigma_j^2}\right] \text{ and } \delta_j = \frac{1}{\left(t-t_{0_j}\right)^2}\left[1 + \frac{\log\left(t-t_{0_j}\right)-\mu_j}{2\sigma_j^2} + \frac{1}{\sigma_j^2}\right]$$

Therefore, by adding the leakage current component, the total AP profile can be expressed by:

$$V(t) \approx A_{Na}\Lambda\left(t; t_{0_{Na}}, \mu_{Na}, \sigma_{Na}\right) - A_K\Lambda\left(t; t_{0_K}, \mu_K, \sigma_K\right) + A_l\,\Lambda\left(t; t_{0_l}, \mu_l, \sigma_l\right) \quad (5)$$

According to the scheme described in Fig.1, an AP is generated when the membrane is depolarized beyond a set threshold, followed by a change in the Na^+ conductance after a delay $t_{0_{Na}}$, where μ_{Na} and σ_{Na} determine the kinematics of opening and closing of the population of Na^+ channels, and A_{Na} is a quantity (in mV.s) proportional to the net influx of Na^+ ions. The same mechanism occurs for the efflux of K^+ ions. According to experimental data, the leakage current flow occurs in the same direction as the Na^+ current.

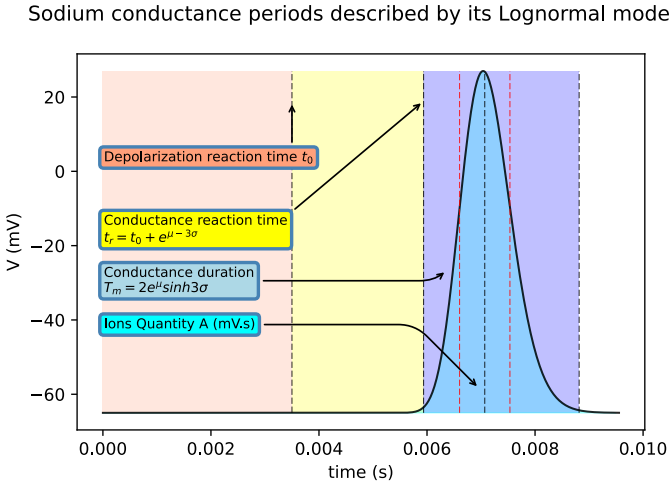


Fig.1: Illustration of the asymmetric bell-shaped Na^+ conductance changes by a lognormal function.

C. *Statistical formulation*

The transmembrane ion transport can be measured as a discrete current randomly switching between two states [24], with the passage of ions observed as stochastic spikes. Experimental results showed that the superposition of hundreds of ionic currents, resulting from a repetitive step depolarization of a channel, leads to an exponential-like profile. This result suggests the ergodicity and causality of ion transport across the membrane, which assumes that the average of a random process parameter over time and the average over the statistical ensemble are the same. In other words, the exponential profile obtained by averaging currents corresponds to the probability density of the passage of ions over channels. By extending this assumption to the membrane behavior, the Na^+ and K^+ AP components can be represented as probability densities, as supported by the HH model and expressed by the Elovich equation. Indeed, Cope [25] showed that the kinetics controlling the transport of Na^+ and K^+ ions across the membrane conforms with the Elovich equation, $^{dq_t}/_{dt} = aexp(-\beta q_t)$, an expression of a concentration decay time q_t of a given ion, which leads to an exponentially decaying of its current. Since a current is proportional to a membrane conductance and represented in the HH model by a probability-like function, the Elovich equation suggests that this current is associated with an ergodic probability density. Considering a neuron membrane comprises many ion channels, the total current measured across the neuron membrane is equal to the sum of the individual ionic currents. In fact, the parallel arrangement of ion channels over the membrane surface and the stochastic nature of ion crossing suggests that the observed conductance profile is a direct calculation of their probability density. As the channel ionic currents decay exponentially, their sum corresponds to a convolution product of densities related to individual currents. For a large number of channels, using the central limit theorem, this pattern tends towards a lognormal limit profile [26,27].

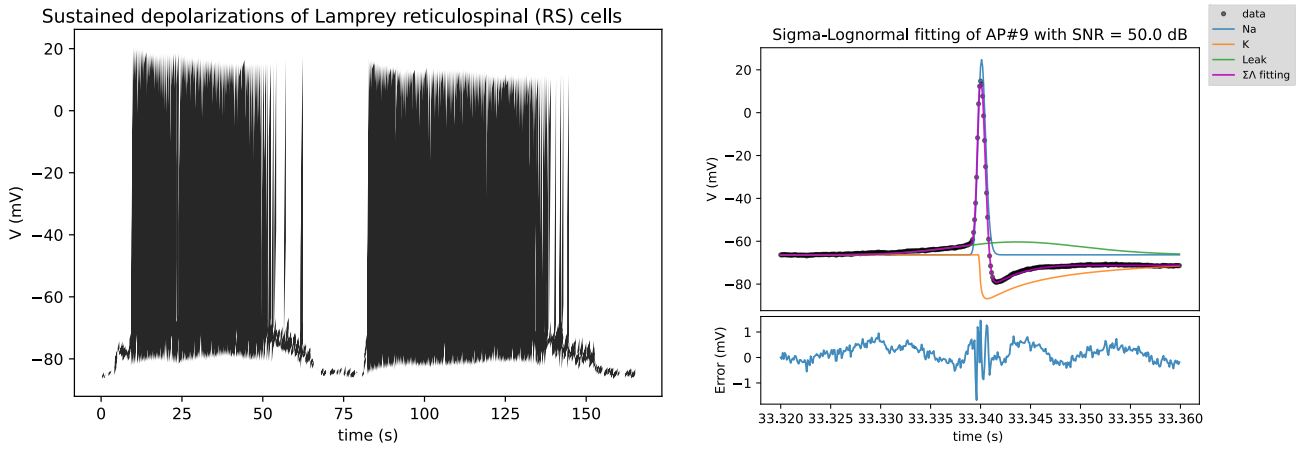
III. FITTING CONSIDERATIONS

A. *DATA DESCRIPTION*

For this study, we used AP recordings from a previous study [28]. In short, the data were recorded during depolarizations of lamprey reticulospinal (RS) cells, sustained for up to minutes by sensory stimulations. Different modes of stimulation (mechanical pressure on the head skin, electrical shocks, etc.) were applied to elicit the depolarizations. The experiments were performed on an *in vitro* semi-intact preparation. The recorded RS cells were located in the middle rhomb encephalic reticular nuclei on both sides of the brain. Specifically, single and paired intracellular recordings were performed on the giant Müller cells using Axon instruments (Axoclamp 2A amplifier, Digidata 1322 interface, Clampex 9.2 digitizer software) and AP waveforms were extracted (Figure 2.a). The original biosignal, described in Fig.2-a, is segmented into windows of 50 ms width, centered on the AP peak. This segmentation led to 1046 APs, which we fitted with the Sigma-Lognormal equation (5). A typical fitting result, as presented in Fig.2-b, is summarized by the following lognormal parameters of the sodium and potassium components:

	Sodium component				Potassium component			
Parameters / AP	μ	σ	A (mV.s)	t_0(s)	μ	σ	A (mV.s)	t_0(s)
AP#9	-5.624	0.106	0.087	33.337	-3.322	1.954	0.536	33.340

B. FITTING RESULTS AND INTERPRETATIONS



(a) (b)

Figure 2. a) Example of raw data showing RS cell response. b) Example of typical fitting result of AP, extracted from (a), with the Sigma-Lognormal model.

The raising of AP can be described as follows. It occurred after various input signals were received by neuron dendrites, increasing its membrane potential, which reached its depolarization threshold at $t_{0_{Na}} = 33.337$ s, after the beginning of the recording. This time underlines the existence of delay between the time when the depolarization is launched and when the inward flow of sodium ions arises, i.e., the conductance starts. Another interesting benefit of the Sigma-Lognormal model is its ability to distinguish contributions of sodium and potassium conductance in the AP waveform, a similarity with the method used in Voltage-Clamp experiments by Hodgkin et Huxley to isolate sodium and potassium currents, which led to their mathematical model. Accordingly, specific characteristic times are computed to describe the progression of the depolarization effect and the emergence of the AP waveform. For example, as depicted in the above results, the K^+ conductance, triggered by the Na^+ influx at $t_{0_K} \approx 33.340$ s occurred 3 ms after $t_{0_{Na}}$, which represents approximately the time when the Na^+ activity reached its maximum change $t_{inflex_{Na}} = e^{\mu-\sigma^2-0.5\sigma(\sigma+\sqrt{4+\sigma^2})} \approx 3.2\ ms$. This delay is expected from the biophysics of the AP as these ion channels start opening in mass only when the Na^+ influx has depolarized the membrane. Furthermore, as sodium and potassium conductance can be analyzed separately, one can evaluate their time duration through the AP waveforms. Indeed, from Fig.1, increased conductance durations $T = 2e^{\mu}\sinh 3\sigma$ are respectively equal to $T_{Na} \approx 2.3\ ms$ and $T_K \approx 12.7\ ms$. This longer duration for the K^+ process is also expected by biophysics observations, as its conductance waveform obtained in Voltage-Clamp experiment has been reported in the fourth original HH paper [16] to be temporally broader than the sodium one waveform. Such interpretation complements what is currently used in this field and provides a new microscopic window into these processes. For example, the evolution of parameter values can provide indications of membrane fatigue.

IV. COMPARING THE SIGMA-LOGNORMAL EQUATION WITH THE HH MODEL

Concerning the comparison between this kinematic approach and the Hodgkin-Huxley (HH) original model, historically, this latter was developed to describe ionic currents measured in voltage-clamp experiments, not in the AP waveform itself. The HH model was proposed to explain the ionic mechanisms underlying AP generation, not to fit individual experimental AP waveforms. In other words, the HH model is calibrated by tuning a set of parameters to reproduce the overall behavior of an AP that matches the general experimental profile, but it was not designed for direct data fitting on a point-by-point basis. It is viewed as a generative model of APs, whereas the Sigma-Lognormal model, as presented here, is primarily an analytical operational model dealing with experimental AP data by using a reduced number of parameters that indirectly quantify the dynamics of ions crossing the cell membrane after depolarizations and the temporal sequences of sodium and potassium ionic components involved in shaping the AP waveform. The approach presented in this abstract should thus be seen as complementary to the HH model. It does not attempt to describe the underlying ionic mechanisms generating the AP but rather to quantify some physical parameters, as mentioned above, based on experimental AP recordings.

V. PERTINENCE OF USING SIGMA-LOGNORMAL EQUATION TO MODEL ACTION POTENTIAL WAVEFORMS

The Sigma-Lognormal model has been used to fit the action potential waveforms as the lognormal function satisfies the conditions of equation (4) proposed by Cope [25] to model the hidden dynamics that lead to its stereotyped profile. The approach of using the Kinematic Theory and its model is justified by the fact that the AP arises from the flow of ions triggered by the membrane depolar-

ization, where ionic currents represent charge velocities. Its final shape is made of a superposition of stereotyped bell-shaped components of opposite polarity; a phenomenon explained in the Kinematic Theory by the agonist-antagonist synergy. This computational fitting solution serves to explore new insights through parameters, such as the temporal sequence of transmembrane ionic passages leading to the formation of the action potential (depolarization and repolarization periods). On a conceptual level, in graphonomics, handwriting is analyzed through the study of hand movement kinematics (speed, pressure, trajectory, letter shapes, etc.). The action potential can be analyzed from a similar perspective: its stereotyped profile emerges from the coordinated movement of ions whose bell-shaped temporal profiles share behavioral similarities with velocities ones, including the same underlying mechanisms of agonist-antagonist synergy (sodium and potassium flows).

VI. Conclusion

This paper adds AP on the increasing list of lognormal phenomena observed in nature. Using the Kinematic Theory of movement and its Sigma-Lognormal equation to model its profile open a window for analysing bioelectrical activity measured *in vivo*. The lognormal parameters give more details on the timing of ions activities involved in the formation of the AP and make it possible to retrieve the time course of the membrane conductance changes. Using the Sigma-Lognormal model to describe neuron activity further opens potential applications in numerous fields like physiology, pathophysiology, neuroscience, cardiology, pharmacology, biotechnology as well as for biological and *in silico* neural network analysis and synthesis. To go one step further in these directions, we will next focus on finding a way to get the corresponding number of ions per unit of membrane conductance that have crossed the membrane from the quantities A_{Na}, A_K and A_{Leak}, expressed in volt-seconds to provide new insights on neuron excitability [29].

References

[1] Aitchison, J., Brown, J.A.C. The lognormal distribution, Cambridge University Press 1966.

[2] Crow, E.L., Shimizu, K. Lognormal distributions Theory and Applications, CRC Press, 1988

[3] Limpert, Eckhard & Stahel, Werner & Abbt , Markus. Log-Normal Distributions Across the Sciences: Keys and Clues. BioScience. 51. 341, 2. 2001

[4] Buzsáki G, Mizuseki K. "The log-dynamic brain: how skewed distributions affect network operations". Nat Rev Neurosci. 2014 Apr;15(4):264-78.

[5] Plamondon, R. "A kinematic theory of rapid human movements : Part I: Movement representation and generation". Biological Cybernetics..72:295-307, 1995

[6] Plamondon, R. "A kinematic theory of rapid human movements : Part II. Movement time and control". Biological Cybernetics. 72:309-320, 1995

[7] Djioua, M. Plamondon R "The limit profile of a rapid movement velocity", Human Movement Science (2009).

[8] Plamondon, R.,et al. "A kinematic theory of rapid human movement. Part IV: a formal mathematical proof and new insights". *Biol. Cybern.* 89, 126–138 2003.

[9] Plamondon, R., Marcelli, A., & Ferrer, M. A. (Eds.). The Lognormality Principle and its Applications in e-Security, e-Learning and e-Health, World Scientific Publishing Series in Machine Perception and Artificial Intelligence (88): 1-39, 2021

[10] Plamondon, R., Bensalah, A., Lebel, K., Salameh, R., Séguin de Broin G., O'Reilly, C., Begon, M., Desbiens, O., Guy, A., Fol Lemarie, F, Boyoguéno-Bidias, S-P., FIischer, A., Zhang, Z., Alamargot, D., Rémi, C., Faci, N., Fortin, R., Simard, M-N., Bazinet, C. (2023). Lognormality: an open window on neuromotor control. Springer Lecture Notes in Computer Sciences. Graphonomics in Human Body Movement: Bridging Research and Practice from Motor Control to Handwriting Analysis and Recognition. Proc, 21st In. Conf, of the International Graphonomics Society, IGS 2023., Evora, Portugal (1205-258).

[11] O'Reilly C, Oruganti SDR, Tilwani D, Bradshaw J. Model-Driven Analysis of ECG Using Reinforcement Learning. *Bioengineering*. 2023; 10(6):696.

[12] Djioua, M. Contributions à la compréhension, à la généralisation et à l'utilisation de la théorie cinématique dans l'analyse et la synthèse du mouvement humain, 2007, [Ph.D. thesis, École Polytechnique de Montréal]. https://publications.polymtl.ca/7928/.

[13] Hodgkin A.L., Huxley,A.F. "Currents carried by sodium and potassium ions through the membrane of the giant axon of Loligo", J. Physiology, vol. 116, pp. 449-472., 1952

[14] Hodgkin A.L., Huxley,A.F. The components of membrane conductance in the giant axon of Loligo, J. Physiology, vol. 116, pp. 473-496., 1952

[15] Hodgkin A.L., Huxley,A.F. The dual effect of membrane potential on sodium conductance in the giant axon of Loligo, J. Physiology,vol.116, p.497-506., 1952

[16] Hodgkin A.L., Huxley,A.F. "A quantitative description of membrane current and its application to conduction and excitation in nerve", J. Physiology, vol. 117, pp. 500-544., 1952

[17] Hodgkin A.L., Huxley,A.F., Katz, B., "Measurement of current-voltage relations in the giant axon of Loligo", J. Physiology, vol. 116, pp. 424-448., 1952

[18] Cole K.S. and Moore J.W. "Potassium ion current in the squid giant axon: dynamic characteristic". Biophysical Journal, 1, pp 1-14. , 1960

[19] Hoyt,R.C. "The Squid Giant Axon. Mathematical Models", Biophysical Journal, vol. 3, pp. 399-431., 1963

[20] Sehgal S, Patel ND, Malik A, Roop PS, Trew ML (2019) Resonant model—A new paradigm for modeling an action potential of biological cells. 14(5):

[21] Clay J.R. and DeFelice L.J. "Relationship between membrane excitability and single channel open-close kinetics", Biophysics Journal, 42, pp 151-157. 1983

[22] Goldman, D.E. A molecular structural basis for the excitation properties of axons, Biophysical Journal vol. 4, pp. 167-188. 1964

[23] Hoyt, R.C. Sodium inactivation in nerve fibers, Biophysical Journal, vol.8, pp. 1074-1097. 1968

[24] Sigworth, F. J., and E. Neher.. Single Na+ channel currents observed in cultured rat muscle cells. Nature (Lond.). 287:447-449. 1980

[25] Cope F.W. A theory of ion transport across cell surfaces by a process analogous to electron transport across liquid-solid interfaces. Bulletin of the Mathematical Biophysics, 27, 99-109. 1965

[26] Papoulis, A. The Fourier Integral and Its Applications. McGraw-Hill. 1962

[27] Thorin, O. On the infinite divisibly of the lognormal distribution," Scandinavian Actuarial Journal. 121-148. 1977

[28] Antri M., Fenelon K. and Dubuc R. The Contribution of Synaptic Inputs to Sustained Depolarizations in Reticulospinal Neurons. Journal of Neuroscience, 29(4) 1140-1151. 2009

[29] Drukarch B, Wilhelmus MMM. Thinking about the action potential: the nerve signal as a window to the physical principles guiding neuronal excitability. Front Cell Neurosci. 2023 Aug 28;17:1232020. doi: 10.3389/fncel.2023.1232020. PMID: 37701723; PMCID: PMC10493309.

Using the Kinematic Theory to Infer a New Optimal Control Cost Function

Najoua Assila
Chair of Psychology of Learning and Instruction,
Centre for Tactile Internet with Human-in-the-Loop
Technische Universität Dresden
Dresden, Germany
najoua.assila@tu-dresden.de

Ben Braithwaite
Département de Génie électrique
Polytechnique Montréal
Montréal, QC, Canada
ben.braithwaite@polymtl.ca

Mickaël Begon
School of Kinesiology and Exercise Sciences
University of Montréal,
Centre de recherche Azrieli du CHU Sainte-Justine
Montréal, QC, Canada
mickael.begon@umontreal.ca

Réjean Plamondon
Département de Génie électrique
Polytechnique Montréal
Montréal, QC, Canada
rejean.plamondon@polymtl.ca

***Abstract*— Velocity profiles of rapid human movement are strongly stereotypical. Understanding their invariants is a key for understanding motor control. The Kinematic Theory links these invariants to the structure of the neuromuscular system, while the optimal control theory considers that these invariants result from the optimality of the control strategy. Although these two interpretations are fundamentally different, they are not opposed. Additionally, as the lognormal impulse response predicted by the Kinematic Theory produces the most realistic velocity profiles, we aimed to investigate this function from an optimality perspective. Through this analysis, we introduce a new optimality criterion, that we used to predict the control strategy of a simplified arm extension. Our new composite cost function produced asymmetric velocity profiles, that are the closest to those predicted using the Kinematic Theory, furthering our understanding of motor control strategies.**

Keywords*— *Kinematic Theory, optimal control, fast movement, motor control strategies

I. Introduction

Investigating the stereotypical patterns of healthy human kinematics offers insights into the understanding of movement disorders and the development of targeted rehabilitation strategies. The emergence of these kinematic invariances has been interpreted through different theoretical frameworks. For instance, the Kinematic Theory of rapid human movements attributes velocity invariants to the coupling of subsystems within the neuromuscular system, postulating that these subsystems cumulative delays are not independent, but obey a proportionality relationship [1]. As a result, the neuromuscular system's impulse response converges toward a lognormal function as the number of subsystems increases [2]. Another perspective comes from the optimal control framework, which suggests that the neuromuscular system operates optimally, with control strategies reflecting this optimality under given constraints [3]. However, the specific optimality criterion governing human movements remains uncertain, with insights provided from experimental observations. Since the velocity profiles predicted by the Kinematic Theory align most closely with empirical data, we propose investigating its lognormal function from an optimality perspective. This theoretical analysis aims to derive a new optimality criterion capable of predicting velocity profile invariants (*e.g.*, the asymmetry of the velocity profile). The criterion will be evaluated using a simplified arm extension model.

II. Cost Function Development

A. Commonly used control strategies

Human movement trajectory are planned using the neuromusculoskeletal kinematics characteristics [4]. Accordingly, these characteristics are strong candidates to integrate within optimal control cost functions that express potential motor control strategies. Indeed, the norm of different order derivatives of the position variables have previously been investigated [5]–[7]. Taken together, these studies indicate that, compared to other cost functions, minimizing jerk (*i.e.*, third-order derivative) produces the most realistic movement [8]. Nevertheless, they do not exclude the possibility that the cost function could be a combination of different order derivatives. Indeed, Ben-Itzhak and Karniel [9] highlighted that movement predicted through a combination of jerk and acceleration were superior to those predicted solely through minimizing jerk. Interestingly, these

This research was conducted as part of the TransMedTech Institute's activities, and Ben Braithwaite was funded in part by the Canada First Research Excellence Fund. This work was also partly supported by NSERC Discovery Grants RGPIN-2015-06409 to Réjean Plamondon and RGPIN-2019-04978 to Mickaël Begon. Najoua Assila was partially funded by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany's Excellence Strategy – EXC 2050/1 – Project ID 390696704 – Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop" (CeTI) of Technische Universität Dresden.

cost functions set the movement time as a fixed boundary condition. However, the movement duration is the result of a trade-off between efficiency and expediency [10], potentially explaining why these cost function fail to capture some invariant characteristics of human movement, *e.g.*, the asymmetry of the velocity profile [11]. Indeed, a minimum-time model, supposing a "bang-bang" control, predicts a Gaussian velocity profile [11], [12], suggesting that a time-optimal control strategy is critical to produce an asymmetric velocity profile.

B. Lognormal parameters effects on the control strategies

The Kinematic Theory predicts that the velocity profile is defined by a lognormal function:

$$v(t) = D\Lambda(t, t_0, \mu, \sigma^2)$$
$$\text{where} \quad \Lambda(t, t_0, \mu, \sigma^2) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}(t-t_0)} \exp(-\frac{(\ln(t-t_0)-\mu)^2}{2\sigma^2}) & if \;\; t_0 < t, \\ 0 & otherwise \end{cases} \tag{1}$$

with D the amplitude of the motor command, t_0 the time occurrence of the motor command, μ the logtime delay and σ the logresponse. The logtime parameters (μ and σ) modulate the shape of the velocity profile. In particular, σ controls the asymmetry of the system, while μ describes its underlying activation dynamics.

Using the Kinematic Theory definition of the velocity profile, we can define the acceleration and jerk profiles, and evaluate their respective norm-based cost-functions. To assess how the lognormal parameters influence the cost associated with each derivative models, we compute the 2-norm of the n-th order derivative of a position variable x as a function of $\mu \in [-3.0, -1.5]$ and $\sigma \in [0.1, 0.6]$ values for arm fast movements [13]–[15].

$$N_2(n, \mu, \sigma) = \sqrt{\int_{t_{start}}^{t_{end}} \left(\frac{d^n x}{dt^n}\right)^2 dt} \tag{2}$$

Irrespective of the derivative order, the cost function decreases with the increase of μ and σ, although the decrease is faster with σ (Fig. 1). Accordingly, using any of these cost functions, the optimal movement is equivalent to maximizing the lognormal parameters. From a Kinematic Theory perspective, this increase would indicate that the neuromuscular system constitutive subsystems are getting slower and the delay between their successive activation is increasing. This activation dynamic describes a motor strategy incompatible with the execution of a fast movement. Additionally, no physiologically-sound linear combination of these cost functions is sufficient to solve this problem. Thus, one must consider introducing additional terms to the cost function.

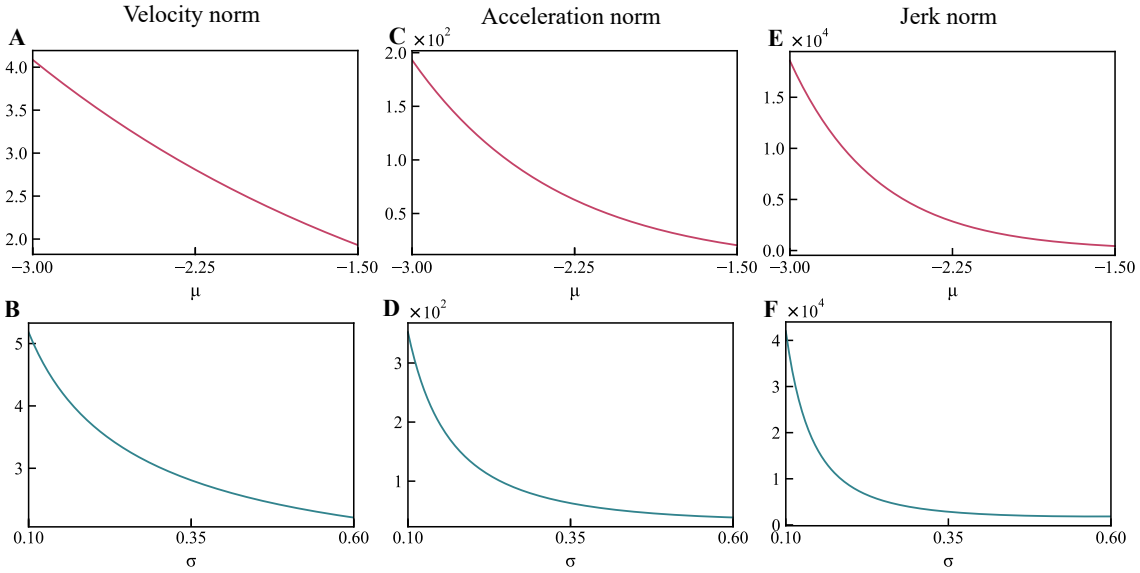


Fig. 1. The 2-norm velocity (A and B), acceleration (C and D) and jerk (E and F) as a function of the lognormal parameters (top: μ with $\sigma = 0.35$ and bottom: σ with $\mu = -2.25$).

C. Fast movement as a trade-off between its duration, energy and smoothness

The analysis of the previously introduced cost functions highlights the impossibility of either using a single cost function or using only kinematics-based cost function to optimally control human movements. A nontrivial optimal control strategy (*i.e.*, neither extremely slow nor unrealistically fast) appears when there is a trade-off between opposing objectives (*e.g.*, the movement smoothness and its duration). Additionally, Berret et al. [16] have shown that kinetic energy, as an indicator of the mechanical energy cost, is another key factor. All these considerations lead us to hypothesize that the neuromuscular system motor strategies emerge as a compromise between movement smoothness, energy efficiency and movement time. Accordingly, we introduce the composite cost function:

$$C_{OCP}(\mathbf{q}, w) = \int_{t_{start}}^{t_{end}} [w \times \mathbf{v}^2(\mathbf{q}, t) + \mathbf{j}^2(\mathbf{q}, t)] \times (t - t_0)^2 dt \tag{3}$$

with $\mathbf{q}$ a positional vector, $\mathbf{v}$ its first order derivative (*i.e.*, velocity), $\mathbf{j}$ its third order derivative (*i.e.*, jerk), and w a weighting factor.

III. Applying the composite function to predict arm extension

A. Problem definition

We studied a simplified forearm movement, representative of a movement used to draw a curved line. The arm is modelled with a single degree of freedom that represents a rotation in the horizontal plane. The forearm segment is 30 cm long and weighs 2.4 kg. To evaluate the segmental moments of inertia, the segment was simplified as a rigid cylinder with a radius of 3 cm and a uniformly distributed mass. The upper and lower bounds for velocity, acceleration and jerk were defined based on previously reported experimental data of fast reaching movements [17].

We simulated a fast extension constraining the elbow angle (θ) at the start and end of the movement. We predicted optimal velocity profiles with regard to the jerk (C_j), the jerk and time (*i.e.*, $w = 0$ in (3); C_{jt}); or the jerk, time and energy (C_{OCP}). We set $t_0 = 0.01\,s$ and $w = 5000$ in (3), to balance the contributions of the two costs, as the jerk's magnitude is larger than the velocity's. No kinetic constraints were imposed, as we limited our analysis to the kinematic outputs. Thus, we opted to use the jerk as a control variable, which also improved the dynamic consistency [18]. The complete problem can be written as:

$$\begin{aligned} &\min_{\mathbf{q},u} C_{OCP}(\mathbf{q}, w) \\ &\mathbf{q} = (\theta, \dot{\theta}), u \equiv \dddot{\theta}, w = 5000 \\ s.t. \quad \forall t \quad &(\theta(t) \in [-110^\circ, 30^\circ],\ |\dot{\theta}| < 600^\circ/\mathrm{s},\ |\ddot{\theta}| < 5 \times 10^{3\circ}/\mathrm{s}^2,\ |\dddot{\theta}| < 4 \times 10^{3\circ}/\mathrm{s}^3) \\ &\mathbf{q}(t_{start} = 0\,\mathrm{s}) = (30^\circ, 0^\circ/\mathrm{s}) \quad \text{and} \quad \mathbf{q}(t_{end} = 8\,\mathrm{s}) = (-70^\circ, 0^\circ/\mathrm{s}) \end{aligned} \tag{4}$$

To solve the optimal control problem numerically, we use a direct multiple shooting transcription with 60 shooting points and a 4$^{\text{th}}$-order *Runge-Kutta* integration with 5 intermediate steps [19]. The nonlinear program was solved using *Ipopt* [20].

B. Analysis

A lognormal function was fitted to each of the profiles predicted by the cost functions (C_j), (C_{jt}) and (C_{OCP}) [15], while constraining t_0 to its defined value in the optimal control problem, and σ to its physiological range. The similarity between the velocity profiles resulting from the optimal control problem and their respective fitted lognormal function was assessed by calculating a Signal-to-Noise Ratio (SNR). As velocity profiles predicted by the Kinematic Theory align best with experimental observations, we consider the signal to be the lognormal velocity profile, and the noise to be the difference between the lognormal and the simulated velocity profiles. Accordingly, the higher the SNR, the better the fit between the optimal control velocity prediction and the lognormal profile.

C. Results and discussion

The predicted kinematic profiles differ depending on the cost-function (Fig. 2). Fitting a lognormal to each profile leads to different lognormal parameters and SNR. Indeed, the velocity profiles of (C_j), (C_{jt}) and (C_{OCP}) are the closest to lognormal profiles with parameters (μ, σ) of (-0.7, 0.45), (-1.12, 0.6) and (-1.42, 0.6), respectively. Accordingly, the SNR for the minimum jerk, minimum jerk and time, and composite models were 10.51, 13.76, and 14.09 dB.

Minimizing jerk (C_j) produces an unrealistic symmetrical velocity profile, as expected [8], [12], [21], which explains its low SNR. In contrast, minimizing both jerk and time (C_{jt}) or jerk, time and energy (C_{OCP}) leads to asymmetrical velocity profiles, *i.e.*, skewed towards the beginning of the movement. This aligns with the literature, as the minimum-time model predicts asymmetric velocity profiles [12]. Moreover, comparing the jerk-time cost function (C_{jt}) to the composite function (C_{OCP}), the latter's velocity profile shows a more concave deceleration phase and a faster acceleration. These profile changes

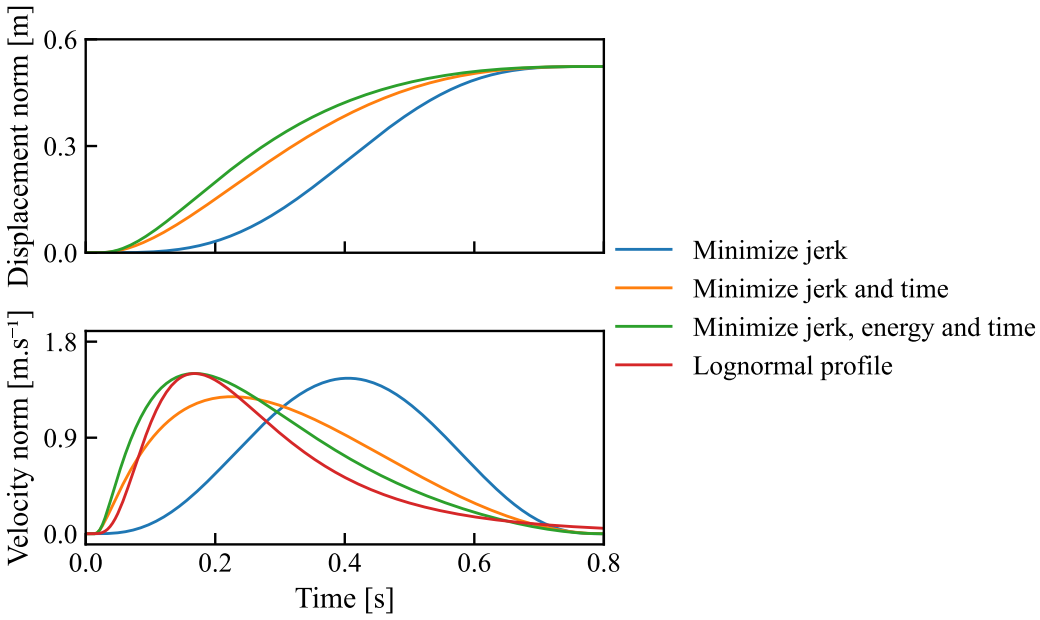


Fig. 2. Displacement and velocity profiles predicted by different objective functions: minimum jerk (C_j, blue); minimum jerk and time (C_{jt}, orange); and the composite function (C_{OCP}, green) with $w = 5000$. A lognormal velocity profile ($\mu = -1.42$, $\sigma = 0.6$) was fitted to the composite function (red).

lead to a higher SNR, as well as a logtime delay closer to the physiological range. This suggests that the composite function may provide more physiologically plausible predictions.

While our composite function produces an inflection in the deceleration phase, its predicted velocity remains convex during the acceleration phase. Indeed, the weight introduced by the time renders the objective function almost null at the start of the movement regardless of the predicted jerk or velocity values. Thus, the problem is minimally constrained by the optimality criterion at the beginning of the movement. From a mechanical perspective, this indicates that the current optimal control problem fails to account for the difficulty of countering the inertia of the initially static arm. The use of a dynamic model with joint torque constraints may better express inertial effects, thereby adding implicit structural constraints to the optimal control.

According to the Kinematic Theory, any dynamical system that can be modeled by convolving an infinite number of proportionally coupled subsystems will have a lognormal impulse response. The resulting lognormal velocity profiles have been widely used to characterize wrist, hand, and arm movements across various experimental conditions, in both healthy individuals and those with neurological disorders such as Parkinson's and Alzheimer's [22], [23]. Analytical models based on minimum principles have been shown to approximate these lognormal profiles to varying degrees of accuracy [24], positioning the Kinematic Theory as an ultimate minimization framework. Building on this perspective, we employed the lognormal profile to investigate the limitations of traditional kinematic optimization criteria in predicting realistic velocity profiles. Our findings suggest that these criteria often fail to account for the temporal constraints inherent to rapid movements. Using a simple single-degree-of-freedom model, we demonstrated that a cost function incorporating jerk, energy, and time can closely match the deceleration phase of movement, though it falls short in reproducing the initial acceleration. It could be argued that the application of optimal control with a multi-criterion control strategy, involving musculoskeletal parameters, to plan motion would give more realistic results. However, these constraints are not necessary in the Kinematic domain when lognormal impulse response are used. Thus, more realistic velocity profiles may be achievable within the kinematic control space alone, provided that the optimal control criterion effectively captures the motor control system's underlying organization, central to the Kinematic Theory.

References

- R. Plamondon, "A kinematic theory of rapid human movements: Part I. Movement representation and generation," Biol. Cybern., vol. 72, no. 4, pp. 295–307, Mar. 1995, doi: 10.1007/BF00202785."
- R. Plamondon, C. Feng, and A. Woch, "A kinematic theory of rapid human movement. Part IV: a formal mathematical proof and new insights," Biol. Cybern., vol. 89, no. 2, pp. 126–138, Aug. 2003, doi: 10.1007/s00422-003-0407-9.
- D. A. Kistemaker, J. D. Wong, and P. L. Gribble, "The cost of moving optimally: kinematic path selection," J. Neurophysiol., vol. 112, no. 8, pp. 1815–1824, Oct. 2014, doi: 10.1152/jn.00291.2014.
- D. M. Wolpert, Z. Ghahramani, and M. I. Jordan, "Are arm trajectories planned in kinematic or dynamic coordinates? An adaptation study," Exp. Brain Res., vol. 103, no. 3, pp. 460–470, Nov. 1995, doi: 10.1007/BF00241505.
- H. Hatze and J. D. Buys, "Energy-optimal controls in the mammalian neuromuscular system," Biol. Cybern., vol. 27, no. 1, pp. 9–20, Mar. 1977, doi: 10.1007/BF00357705.

[6] Y. Uno, M. Kawato, and R. Suzuki, "Formation and control of optimal trajectory in human multijoint arm movement," Biol. Cybern., vol. 61, no. 2, pp. 89–101, Jun. 1989, doi: 10.1007/BF00204593.
[7] D. Eager, A.-M. Pendrill, and N. Reistad, "Beyond velocity and acceleration: jerk, snap and higher derivatives," Eur. J. Phys., vol. 37, no. 6, p. 065008, Oct. 2016, doi: 10.1088/0143-0807/37/6/065008.
[8] M. D. Klein Breteler, R. G. J. Meulenbroek, and S. C. A. M. Gielen, "An evaluation of the minimum-jerk and minimum torque-change principles at the path, trajectory, and movement-cost levels," Motor Control, vol. 6, no. 1, pp. 69–83, Jan. 2002, doi: 10.1123/mcj.6.1.69.
[9] S. Ben-Itzhak and A. Karniel, "Minimum acceleration criterion with constraints implies bang-bang control as an underlying principle for optimal trajectories of arm reaching movements," Neural Comput., vol. 20, no. 3, pp. 779–812, Mar. 2008, doi: 10.1162/neco.2007.12-05-077.
[10] B. Hoff, "A model of duration in normal and perturbed reaching movement," Biol. Cybern., vol. 71, no. 6, pp. 481–488, Oct. 1994, doi: 10.1007/BF00198466.
[11] M. Djioua and R. Plamondon, "The limit profile of a rapid movement velocity," Hum. Mov. Sci., vol. 29, no. 1, pp. 48–61, Feb. 2010, doi: 10.1016/j.humov.2009.02.007.
[12] S. E. Engelbrecht, "Minimum Principles in Motor Control," J. Math. Psychol., vol. 45, no. 3, pp. 497–542, Jun. 2001, doi: 10.1006/jmps.2000.1295.
[13] Z. Pan, S. Talwar, R. Plamondon, and A. W. A. Van Gemmert, "Characteristics of bi-directional unimanual and bimanual drawing movements: The application of the Delta-Lognormal models and Sigma-Lognormal model," Pattern Recognit. Lett., vol. 121, pp. 97–103, Apr. 2019, doi: 10.1016/j.patrec.2018.05.008.
[14] A. Woch, R. Plamondon, and C. O'Reilly, "Kinematic characteristics of bidirectional delta-lognormal primitives in young and older subjects," Hum. Mov. Sci., vol. 30, no. 1, pp. 1–17, Feb. 2011, doi: 10.1016/j.humov.2009.10.006.
[15] M. Djioua and R. Plamondon, "A New Algorithm and System for the Characterization of Handwriting Strokes with Delta-Lognormal Parameters," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 11, pp. 2060–2072, Nov. 2009, doi: 10.1109/TPAMI.2008.264.
[16] B. Berret, E. Chiovetto, F. Nori, and T. Pozzo, "Evidence for Composite Cost Functions in Arm Movement Planning: An Inverse Optimal Control Approach," PLOS Comput. Biol., vol. 7, no. 10, p. e1002183, Oct. 2011, doi: 10.1371/journal.pcbi.1002183.
[17] P. Pigeon, S. B. Bortolami, P. DiZio, and J. R. Lackner, "Coordinated Turn-and-Reach Movements. II. Planning in an External Frame of Reference," J. Neurophysiol., vol. 89, no. 1, pp. 290–303, Jan. 2003, doi: 10.1152/jn.00160.2001.
[18] P. Puchaud, E. Charbonneau, B. Michaud, and M. Begon, "Optimality equivalence and computational advantages of free-floating base dynamics compared to full-body dynamics," Mech. Mach. Theory, vol. 181, p. 105164, Mar. 2023, doi: 10.1016/j.mechmachtheory.2022.105164.
[19] B. Michaud, F. Bailly, E. Charbonneau, A. Ceglia, L. Sanchez, and M. Begon, "Bioptim, a Python Framework for Musculoskeletal Optimal Control in Biomechanics," IEEE Trans. Syst. Man Cybern. Syst., vol. 53, no. 1, pp. 321–332, Jan. 2023, doi: 10.1109/TSMC.2022.3183831.
[20] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," Math. Program., vol. 106, no. 1, pp. 25–57, Mar. 2006, doi: 10.1007/s10107-004-0559-y.
[21] P. Viviani and T. Flash, "Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning," J. Exp. Psychol. Hum. Percept. Perform., vol. 21, no. 1, pp. 32–53, Feb. 1995, doi: 10.1037//0096-1523.21.1.32
[22] R. Plamondon, A. Marcelli, and M. A. Ferrer, (Eds.) The Lognormality Principle And Its Applications In E-security, E-learning And E-health. World Scientific Publishing Series in Machine Perception and Artificial Intelligence, 2021.
[23] R. Plamondon et al., "Lognormality: An Open Window on Neuromotor Control," in Graphonomics in Human Body Movement: Bridging Research and Practice from Motor Control to Handwriting Analysis and Recognition, Lecture Notes in Computer Science, vol. 14194, Springer, 2023, pp. 205–258.
[24] M. Djioua and R. Plamondon, "The limit profile of a rapid movement velocity," Human Movement Science, vol. 29, no. 1, pp. 48–61, Feb. 2010, doi: 10.1016/j.humov.2009.02.007.



Session

Handwriting

High-Efficient Handwritten Text Recognition: A Lightweight CNN-Transformer with Self-Supervised Deconvolutional Decoding

Baha Edine Harrath
Sys Eng Synchromedia Laboratory
baha-edine.harrath.1@ens.etsmtl.ca

Rayane Bencharef
Sys Eng Synchromedia Laboratory
rayane.bencharef.1@ens.etsmtl.ca

Mohammed Hamden
Sys Eng Synchromedia Laboratory
mohammed.hamdan.1@ens.etsmtl.ca

Mohamed Cheriet
Sys Eng Synchromedia Laboratory
mohamed.cheriet@etsmtl.ca

***Abstract*—** Handwritten Text Recognition (HTR) constitutes a fundamental challenge in modern document analysis systems. While significant progress has been made, achieving both high accuracy and computational efficiency remains an open challenge. We present Efficient HTR, a novel architecture that advances the state-of-the-art through three key contributions: First, we introduce a self-supervised learning strategy that combines masked spatial reconstruction with contrastive learning, enabling robust feature extraction without relying on extensive labeled data. Second, we develop an innovative deconvolution-based decoder that efficiently transforms two-dimensional feature maps into one-dimensional sequences, eliminating the need for explicit text line segmentation. Third, we propose a resource-efficient hybrid CNN-transformer architecture that achieves superior accuracy while significantly reducing computational demands. Our comprehensive evaluations demonstrate competitive performance on IAM and RIMES datasets while maintaining lower computational complexity compared to current methods.

Keywords*— *Handwritten Text Recognition, Self-Supervised Learning, Energy-Efficient Deep Learning, Computer Vision

I. INTRODUCTION

Handwritten Text Recognition (HTR) represents a cornerstone challenge in document analysis, bridging handwriting analysis, cognitive science, and computational modeling [1, 2]. The complexity of this task is particularly evident at the paragraph level, where systems must simultaneously handle diverse writing styles, irregular spatial layouts, and intricate structural relationships [3, 4]. Despite significant advances through deep learning, current approaches face three critical limitations: the requirement for extensive labeled datasets [5], dependence on pre-segmented inputs [6, 7], and substantial computational overhead [8, 9].

The evolution of HTR architectures has seen various approaches attempting to address these challenges. Multi-Dimensional Long Short-Term Memory (MDLSTM) models [10, 11] demonstrated strong sequential modeling capabilities but encountered convergence issues and computational bottlenecks. More recent attention-based methods, notably Joint Attention MDLSTM [10] and Vertical Attention Models [15], successfully eliminated the need for explicit segmentation. Transformer-based architectures like LexiconNet [14] further improved accuracy via vocabulary constraints. More recently, DTrOCR [26] employed a decoder-only Transformer to advance paragraph recognition. Similarly, Retsinas et al. [27] demonstrated that integrating MobileViT modules into a CRNN backbone yields competitive results in a more compact form. Kizilirmak & Yanikoglu [25] showed that a pure CNN–BiLSTM pipeline can maintain high accuracy while significantly reducing model complexity. However, these solutions still incur substantial model complexity or rely on predefined lexicons, limiting their deployment in resource-constrained environments.

To overcome these limitations, we propose three primary contributions:

First: We introduce a novel self-supervised learning framework that combines masked spatial reconstruction with contrastive learning [16, 17]. This approach enables robust feature learning directly from paragraph-level inputs without requiring segmentation or extensive labeled data, capturing both fine-grained character details and global structural patterns.

Second: Our architecture incorporates a specialized decoder that efficiently transforms two-dimensional visual features into one-dimensional character sequences while preserving critical spatial relationships [18]. This design enables accurate multi-line text recognition without requiring explicit line segmentation, significantly reducing computational complexity.

Third: We develop a lightweight model that strategically combines convolutional neural networks with Vision Transformers (ViTs) [19]. This hybrid approach achieves state-of-the-art accuracy while maintaining substantially lower computations.

II. THE PROPOSED METHOD

This approach consists of a two-phase framework: a self-supervised pretraining phase that learns contextual representations from unlabeled data, followed by a supervised fine-tuning phase that optimizes these representations for handwritten text recognition (HTR), as illustrated in Figure 1.

A. Network Architecture

The architecture combines efficient CNN blocks for local feature extraction with strategic placement of Vision Transformer (ViT) layers for global context modeling, followed by a deconvolution-based decoder for sequence generation. This design achieves strong performance while maintaining computational efficiency.

a) Feature Extraction and Processing

The network employs CNN blocks based on Howard et al.'s [20] efficient design principles. Each block comprises three key operations:

- **Dimension expansion** through 1×1 convolution.
- **Depth-wise 3×3 convolution** for feature extraction.
- **Linear projection** via 1×1 convolution.

This structure reduces parameters while maintaining feature representation capacity. The initial layers map inputs to a higher-dimensional space where spatial relationships are preserved, crucial for handling the complex layouts of handwritten text.

b) Global Context Modeling

Following Mehta et al. [19], we integrate ViT blocks as residual units to capture long-range dependencies. These blocks process non-overlapping patches through self-attention mechanisms, enabling the model to understand document-level structure. By limiting transformer layers to later stages of the network, we achieve global context modeling while controlling computational cost.

c) Deconvolution-Based Sequence Generation

The decoder employs ConvTranspose2D layers to reconstruct spatial information from compressed features. This approach offers three key advantages:

- **Preservation of spatial** relationships during upsampling.
- **Efficient handling of multi-line text** without explicit segmentation.
- **Smooth transition from 2D visual features to 1D character sequences.**

The final layers map features to character probabilities through a (vocab_size+1) ×42 matrix, with LogSoftmax activation enabling CTC loss computation.

d) Sequence Prediction and CTC Loss

After mapping visual features to character probabilities through the final layers, we employ the Connectionist Temporal Classification (CTC) loss [28] to align the predicted sequences with the ground-truth labels without requiring explicit frame-wise annotation.

Given an input sequence X and a target label sequence l, the CTC loss is defined as:

$$\mathcal{L}\,\mathrm{CTC} = -\log \sum_{\pi \in \mathcal{B}^{-1}(l)} \prod_{t=1}^{T} p(\pi_t \mid X)$$

where:

- π is a valid alignment path,
- B is the collapsing function that removes repeated labels and blank tokens,
- $p\,(\pi t \mid X)$ is the probability of predicting label πt at time t.

This loss formulation enables the model to handle variable-length sequences and flexible alignments, making it particularly suitable for segmentation-free handwritten text recognition.

Table 1: Architectural Specifications

Processing Stage	Layer Configuration	Input Shape → Output Shape	Parameters
Input Processing	Conv2D + CNN Block×2	40×170 → 16×40×170	16 channels
Feature Extraction	MaxPool + Conv2D + CNN Block×2	16×40×170 → 64×20×85	64 channels
Mid-level Processing	MaxPool + Conv2D + CNN Block×2	64×20×85 → 128×10×42	128 channels
Global Modeling	Conv2D + ViT Block×2	128×10×42 → 256×5×21	256 channels
Sequence Generation	ConvTranspose2D + Conv2D	256×5×21 → (V+1) ×42	V = vocab size

This architecture achieves efficient parameter utilization through:

- **Progressive feature compression** (170 → 85 → 42 → 21).
- **Strategic placement of computation-intensive layers** to optimize computational cost.
- **Balanced local and global feature processing** to enhance recognition accuracy.

The combination of these elements enables accurate text recognition while maintaining computational efficiency, as demonstrated by our experimental results.

B. Self-Supervised Pretraining

The pretraining phase focuses on learning robust contextual representations from unannotated handwritten text data via two tasks.

1. Masked Spatial Reconstruction

We employ Masked Spatial Reconstruction to enable the encoder to learn both local character structures and global paragraph layouts through reconstruction of partially occluded images [16]. During training, we randomly mask 30% of input pixels as rectangular blocks, replacing them with zeros or noise. For an input image $I \in R^{H\times W}$, we apply a binary mask $M \in \{0,1\}^{H\times W}$ to generate the masked image: $\mathbf{I_{masked}} = I \odot M$, where $\odot$ denotes element-wise multiplication. The masked image is processed through an encoder comprising CNN blocks followed by ViT blocks, producing a latent representation Z that captures both fine-grained character details and paragraph structures. A reconstruction layer then projects this representation back to pixel space, generating the reconstructed image I_r with values constrained to [0,1] through sigmoid activation. We optimize the reconstruction using Mean Squared Error (MSE)**:** $\mathbf{L_{msr}} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \hat{X}_i)^2$, where X_i and $\hat{X}_i$ represent original and reconstructed pixel values, respectively. This process enables the encoder to learn dependencies by reconstructing missing regions from surrounding context.

2. Contrastive Learning

The contrastive learning component enhances the encoder's robustness to handwriting variations by generating invariant and discriminative representations [17]. We create multiple augmented views of each handwritten paragraph through random transformations including rotation, noise addition, blurring, and contrast adjustments. The encoder processes these augmented views to produce latent embeddings. We optimize these embeddings using the **InfoNCE** loss: $L_c l = -\log\left[\exp\left(sim(z_i, z_j)/\tau\right) / \sum \exp(sim(z_i, z_k)/\tau)\right]$, where z_i and z_j are embeddings of positive pairs, sim(·,·) denotes cosine similarity, τ is a temperature hyperparameter, and the summation spans all samples in the batch. This formulation encourages similarity between positive pairs while maintaining distinction from negative samples.

3. Unified Learning Objective

We combine both tasks into a unified objective function that balances spatial reconstruction with representation learning: $\boldsymbol{L_{ssl}} = \lambda_1 L_{msr} + \lambda_2 L_{const}$, where λ_1 and λ_2 are hyperparameters that control the contribution of each component. This unified approach enables simultaneous learning of local and global structures while developing robust, transformation-invariant representations.

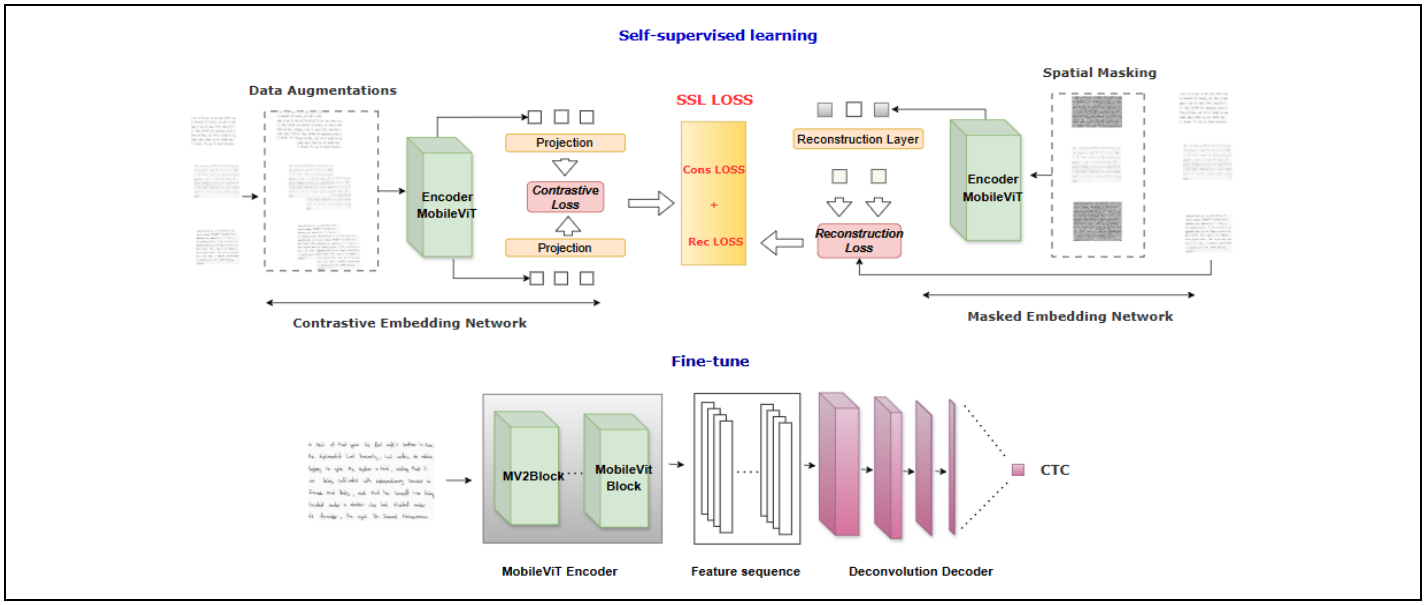


Figure 1. The proposed system consists of **1)** Self-Supervised Learning, where the encoder learns invariant representations via contrastive loss and reconstructs masked regions via spatial masking, optimized with a unified SSL loss. **2)** Fine-Tuning, where the pretrained **MobileViT** encoder extracts features, processed by a deconvolution-based decoder, generating structured text sequences optimized with CTC loss for segmentation-free recognition.

III. EXPERIMENTS AND RESULTS

We conducted comprehensive evaluations of our approach on standard benchmark datasets to assess both recognition accuracy and computational efficiency.

a) Experimental Setup

The evaluation utilizes two widely adopted handwritten text benchmarks: **IAM** [22] and **RIMES** [21]. The IAM dataset contains English handwritten paragraphs from the LOB corpus, comprising 747 training paragraphs, 116 validation paragraphs, and 336 test paragraphs, all at 300 dpi grayscale resolution. The RIMES dataset consists of French handwritten mail correspondence, with 1,400 paragraphs for training, 100 for validation, and 100 for testing, also at 300 dpi grayscale resolution. We employ two primary metrics for evaluation: Character Error Rate (CER) [23] and Word Error Rate (WER) [24]. CER measures character-level accuracy by calculating the ratio of substitutions, deletions, and insertions to the total number of characters. WER quantifies word-level performance using similar operations at the word level.

b) Implementation Details

The implementation incorporates a systematic data augmentation strategy to enhance the encoder's robustness and reduce overfitting. We apply a diverse set of transformations including geometric modifications (rotation ±15°, perspective ±30°), appearance adjustments (brightness ±20%, contrast ±30%), and structural variations (elastic distortions α=40, dilation, erosion). Additionally, we introduce noise (Gaussian σ=0.1, speckle) and blurring effects. Each transformation is applied with 20% probability, ensuring diverse handwriting variations for Contrastive Learning while optimizing InfoNCE loss and improving generalization. The training process consists of two distinct phases:

1. Self-Supervised Pretraining

During pretraining, the model learns contextual representations through Masked Spatial Reconstruction, inferring 30% missing pixels with MSE loss, and Contrastive Learning, optimizing InfoNCE loss on augmented views. These tasks are unified through a weighted loss function with factors $\lambda_1 = 0.7$ and $\lambda_2 = 0.3$, ensuring balanced learning between spatial reconstruction and representation learning.

2. Supervised Fine-Tuning

The pretrained encoder is subsequently adapted for paragraph-level text recognition using labeled data. This phase integrates CNNs for feature extraction, ViTs for global context modeling, and a deconvolution-based decoder for sequence reconstruction. We optimize the final output using CTC loss, with a batch size of 8 and a learning rate of 3e-4 with cosine decay, training for 100 epochs to ensure accurate sequence prediction.

C. Results and Analysis

The experimental results demonstrate that our lightweight model achieves competitive performance on benchmark datasets without requiring external resources or segmentation. The effectiveness of our approach is evidenced by strong performance metrics: on the IAM dataset, we achieve a Character Error Rate of 5.76% and Word Error Rate of 14.23%, while on RIMES, we attain 2.29% CER and 6.88% WER. These results are particularly noteworthy given our model's lightweight architecture and independence from external lexicon data or segmentation requirements. The success of our approach can be attributed to three key factors: 1) the effectiveness of self-supervised pretraining through Masked Spatial Reconstruction and Contrastive Learning, which significantly reduces reliance on labeled data, 2) The balanced design of our hybrid CNN-ViT architecture, which effectively captures both character-level details and global context, and 3) The innovative deconvolution-based decoder, which preserves spatial structure while generating accurate sequences. Table 2 presents a comprehensive comparison with state-of-the-art methods, demonstrating our model's competitive performance while maintaining practical efficiency for real-world applications. Our approach, **DeconvSSL-Net**, achieves comparable or superior results to many existing methods while requiring fewer computational resources and no external lexicons or segmentation.

Table 2: Performance Comparison with State-of-the-Art Methods

Model	IAM Dataset		RIMES Dataset		External Resources
	CER%	WER%	CER%	WER%	Lex/Seg
SAR-MDLSTM+ [11]	7.90	-	-	-	✗/✓
HTR-DT [12]	15.60	-	-	-	✗/✗
JA-MDLSTM [10]	7.90	24.6	2.90	12.6	✗/✗
VAM [15]	4.45	14.55	1.90	6.72	✗/✗
SFR [13]	6.40	23.2	2.10	9.3	✓/✓
LexiconNet [14]	3.24	8.29	1.13	2.94	✓/✗

CNN–BiLSTM[25]	3,59	9,44 %	-	-	✓/✗
DTrOCR [26]	2,38	-	-	-	✗/✓
Enhancing CRNN-HTR[27]	4,22	14,58	2,70	9,46	✗/✗
DeconvSSL-Net (Ours)	**5.76**	**14.23**	**2.29**	**6.88**	✗/✗

IV. Conclusion and Future work

In this paper, we introduced Efficient HTR, a novel self-supervised framework that advances the state-of-the-art in handwritten text recognition through an efficient hybrid architecture and innovative self-supervised learning strategy. Our approach successfully demonstrates that well-designed self-supervised learning, combined with a lightweight CNN-transformer architecture, can achieve competitive recognition performance while reducing computational demands and dependency on labeled data. The experimental results validate our framework's effectiveness, achieving character error rates of 5.76% and 2.29% on IAM and RIMES datasets respectively, without requiring external lexicons or segmentation. An extended version of this work, which will further detail our approach and examine additional datasets to demonstrate the generalization capabilities of our architecture.

References

[1] R. Plamondon and S. N. Srihari, "On-line and off-line handwriting recognition: A comprehensive survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 63–84, 2000.

[2] L. Schomaker, Advances in Handwriting Recognition, Springer, 2007.

[3] T. Bluche, "Joint line segmentation and transcription for end-to-end handwritten paragraph recognition," in Advances in Neural Information Processing Systems (NeurIPS), 2016.

[4] J. Puigcerver, "Are multidimensional recurrent layers really necessary for handwritten text recognition?" in International Conference on Document Analysis and Recognition (ICDAR), 2017, pp. 67–72.

[5] A. Sharma, S. Singh, and S. Ratna, "Graph Neural Network based Handwritten Trajectories Recognition," arXiv preprint arXiv:2405.09247, 2024.

[6] L. Kang, B. Cho, and D. Doermann, "Evaluation of deep convolutional neural networks for document image classification and retrieval," in International Conference on Document Analysis and Recognition (ICDAR), 2014, pp. 186–190.

[7] H. Pham et al., "Robust Handwriting Recognition with Limited and Noisy Data," arXiv preprint arXiv:2008.08148, 2020.

[8] A. Fischer, V. Frinken, and H. Bunke, "Handwriting recognition in historical documents: A survey and analysis," Pattern Recognition, vol. 26, pp. 46–64, 2013.

[9] B. Balci, D. Saadati, and D. Shiferaw, "Handwritten Text Recognition using Deep Learning," CS231n Project Report, Stanford University, 2017.

[10] Bluche T. Joint line segmentation and transcription for end-to-end handwritten paragraph recognition. Adv Neural Inf Process Syst 2016;29.

[11] Bluche T, Louradour J, Messina R. Scan, attend and read: End-to-end handwritten paragraph recognition with mdlstm attention. In: 2017 14th IAPR international conference on document analysis and recognition, Vol. 1. ICDAR, IEEE; 2017, p. 1050–5.

[12] Carbonell M, Mas J, Villegas M, Fornés A, Lladós J. End-to-end handwritten text detection and transcription in full pages. In: 2019 international conference on document analysis and recognition workshops, Vol. 5. ICDARW, IEEE; 2019, p. 29–34.

[13] Wigington C, Tensmeyer C, Davis B, Barrett W, Price B, Cohen S. Start, follow, read: End-to-end full-page handwriting recognition. In: Proceedings of the European conference on computer vision. ECCV, 2018, p. 367–83.

[14] Kumari L, Singh S, Rathore VVS, et al. LexiconNet: An end-to-end handwritten paragraph text recognition system. 2022, arXiv e-prints.

[15] Coquenet D, Chatelain C, Paquet T. End-to-end handwritten paragraph text recognition using a vertical attention network. IEEE Trans Pattern Anal Mach Intell 2022.

[16] X. Chen, S. Xie, and K. He, "An Empirical Study of Masked Autoencoders for Handwritten Document Representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 3, pp. 678–692, 2023.

[17] T. Liu, Z. Lin, and X. Wang, "Contrastive Learning for Document Image Understanding: A Self-Supervised Approach," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11234–11242, 2022.

[18] Y. Zhang, D. Shen, G. Wang, Z. Gan, R. Henao, et L. Carin, « Deconvolutional Paragraph Representation Learning », arXiv preprint arXiv:1708.04729, 2017.

[19] Sachin Mehta and Mohammad Rastegari. Mobilevit: Lightweight, general-purpose, and mobile-friendly vision transformer, 2021.

[20] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, and Hartwig Adam. Searching for mobilenetv3, 2019.

[21] Grosicki, E., El Abed, H.: Icdar 2011-french handwriting recognition competition. pp. 1459–1463 (09 2011).

[22] Marti, U.V., Bunke, H.: The iam-database: An english sentence database for offline handwriting recognition. IJDAR 5, 39–46 (11 2002).

[23] Konstantinidis S. Computing the levenshtein distance of a regular language. In: IEEE information theory workshop, Vol. 2005. IEEE; 2005, p.

[24] Sharma A, Jayagopi DB. Towards efficient unconstrained handwriting recognition using dilated temporal convolution network. Expert Syst Appl 2021;164:114004.

[25] F. Kizilirmak and B. Yanikoglu, "CNN–BiLSTM model for English Handwriting Recognition: Comprehensive Evaluation on the IAM Dataset," arXiv preprint arXiv:2307.00664, 2023.

[26] M. Fujitake, "DTrOCR: Decoder-only Transformer for Optical Character Recognition," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, January 2024.

[27] G. Retsinas, K. Nikolaidou, and G. Sfikas, "Enhancing CRNN HTR architectures with transformer blocks," in Proc. 18th Int. Conf. on Document Analysis and Recognition (ICDAR), Athens, Greece, Aug. 30–Sept. 4, 2024, pp. 425–440.

[28] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks," in Proc. 23rd International Conference on Machine Learning (ICML), Pittsburgh, PA, USA, Jun. 25–29, 2006, pp. 369–376.

Investigating the Effects of Multisensory Integration in Writing and Memorizing New Graphic Patterns

Yoan SEGURA[a], Florence BARA[a], Julien TARDIEU[b], Pierre-Vincent PAUBEL[a] and Jérémy DANNA[a]

[a] *CLLE, University of Toulouse, CNRS, Toulouse, France,* [b] *MSHS-T, University of Toulouse, CNRS, Toulouse, France*

yoan.segura@univ-tlse2.fr, florence.bara@univ-tlse2.fr, julien.tardieu@univ-tlse2.fr, pierre.paubel@univ-tlse2.fr, jeremy.danna@cnrs.fr

***Abstract*— In the digital age, many studies focus on optimizing the learning process of handwriting by modifying the writer's perception of their writing during the action. Among these strategies, it is possible to change the visual perception of the written trace, add auditory feedback with movement sonification, or enhance tactile-kinesthetic feedback using the finger. In this study, the aim was to determine whether there is an optimal method to improve the tracing of new graphic patterns and enhance their memorization. In a first experiment, 32 adult participants learned 16 graphic patterns using a visual and/or auditory strategy, which involved removing the visual feedback of written trace and/or sonifying the handwriting movement. The results indicated a positive effect of removing the written trace on the memorization of graphic patterns, but no effect of sonification on memorization or on the writing process. In a second experiment, the same learning task was presented to another group of 32 adults using an auditory and/or haptic strategy, which involved sonifying the handwriting movement and/or writing with the index finger, while retaining the beneficial visual strategy previously revealed. The results suggested that using the finger aids in memorizing the graphic patterns, while the movement sonification has a negative impact. Neither experiment revealed a significant effect of multisensory integration resulting from the combined strategies. These results encourage further reflection on the relevant strategy of movement sonification and multisensory integration in light of theories of cognitive load.**

Keywords*— *Sensory feedback, Augmented reality, Digital, Motor learning, Visuomotor representation

I. Introduction

The growing interest in new technologies, particularly digital tablets, offers the prospect of optimizing the process of learning to write [1]. The educational benefits are important, given that 10 to 30% of children have handwriting difficulties at school [2]. Currently, the literature reports three main sensory strategies—visual, auditory, and haptic—to optimize this learning process (for a review, see [1]). For instance, a visual strategy consists in reducing the visual perception of the written trace by removing it. This strategy revealed an improvement in movement velocity and fluency in children in letter production, but with a lower precision if the written trace is not perceived after it has been produced [3]. Another strategy consists in applying real-time auditory feedback by associating sound information with the information underlying the handwriting movement. A pseudo-experiment in which children with dysgraphia were trained with this strategy suggested an improvement in movement velocity and fluency in a protocol of rehabilitation, which must be confirmed with a control condition [4]. Finally, a haptic strategy consists in increasing tactile-kinesthetic feedback, using the finger rather than a pen. A study in which adults were trained in this strategy revealed better memorization of graphic patterns traced with the index finger compared to those traced with the stylus [5]. It is worth noting that these sensory strategies share the same objective: sonifying the handwriting movement or writing with the finger reinforces the perception of the handwriting movement, and removing visual feedback of written trace shifts attentional focus from the visual graphic pattern to handwriting movement. Surprisingly, to our knowledge, no study has combined them. Increasing auditory and haptic feedback of the writing process while reducing visual feedback of the writing product could enhance multisensory integration centered on the handwriting movement. Thus, this could reinforce multimodal memory traces and new neural connections between visual, auditory and sensorimotor cortices [6]. This multisensory benefit has already been revealed in the memorization and identification of natural objects [7], and there is every reason to hypothesize it also facilitates the memorization of letters. The aim of this study was therefore to determine whether there is an optimal method for improving the tracing of new graphic patterns and their memorization. Two experiments were conducted. In the first experiment, we tested visual and auditory strategies. In the second experiment, we tested auditory and haptic strategies with the best visual condition resulting from the first experiment.

II. Method of Experiment 1

A. Participants

Thirty-two adult participants took part in the experiment (mean age: 23.53 ± 3.83, 24 female, and 29 right-handed). None of them had any known intellectual, sensorimotor, attentional, affective, or writing deficits. All had normal or corrected-to-normal vision. Prior to the experiment, they signed an informed consent form following the recommendations of the Toulouse Research Ethics Committee (CER 2023_753).

B. Task and procedure

The task consisted in learning 16 new graphic patterns (Fig. 1) [8] by tracing them with the non-dominant hand on a touch screen laid flat. The participants were equipped with binaural headphones and a stylus. The procedure was as follows (Fig. 2):

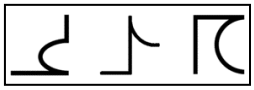


Fig 1. Examples of patterns.

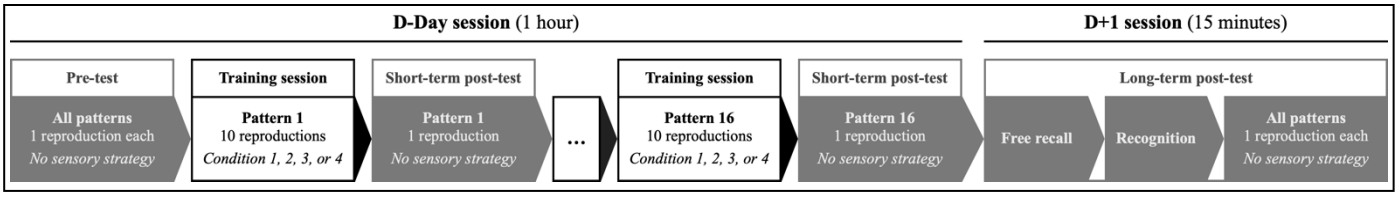


Fig 2. Experimental design of Experiment 1.

Pre-test. After a familiarization phase with the material, the 16 graphic patterns were presented successively on the screen. Participants had to trace each pattern once inside an empty frame below it.

Training session. The patterns were once again presented successively. Participants had to trace each one ten times. These patterns were evenly distributed across four writing conditions according to the presence or absence of visual and auditory strategies (Table I):

- Audiovisual: without written trace and with movement sonification.
- Visual: without written trace and without sonification.
- Auditory: with written trace and with movement sonification.
- Control: with written trace and without sonification (i.e., without sensory strategy).

TABLE I. WRITING CONDITIONS OF EXPERIMENT 1.

Condition	Vision	Sound	Strategie
1	No	Yes	Audiovisual
2	No	No	Visual
3	Yes	Yes	Auditory
4	Yes	No	Control

The conditions were counterbalanced and the patterns were presented in a random order between participants. There was no time limit or speed constraint. During no-trace conditions, the written trace was invisible during reproduction and displayed for 3 seconds after completion. Furthermore, the movement sonification was inspired by the SoundScript method [9]: it varied according to the position of the stylus, in such a way that the pitch varied along the vertical axis (y-axis) and that left-right volume variation in binaural headphones varied along the horizontal axis (x-axis).

Short-term post-test. After the 10 trials, the pattern was traced once in the control condition (i.e., without sensory strategy).

Long-term post-test. First, a free recall task was presented. Participants had to trace from memory as many patterns as they had learned the previous day. Then, a recognition task was presented. Participants had to recognize the 16 learned patterns among 32 distractors—mirror and incorrect stroke—distributed on recognition sheets. Finally, as in the pre-test, the 16 patterns were presented successively and participants had to trace them once.

C. Data analysis

Two aspects were evaluated: writing performance and memorization. Writing performance included mean velocity (in mm/s; from the beginning to the end of the reproduction), mean trace length (in mm), mean movement dysfluency (i.e., the number of abnormal velocity peaks, measured by the SNvpd method [10]), and mean spatial error (i.e., the spatial difference between production and the model; for more details, see [11]). Memorization was evaluated by calculating the number of graphic patterns correctly recalled/recognized divided by the number of patterns learned in each condition (as a ratio between 0 and 1).

All statistical analyses were conducted using JAMOVI® and RStudio®, with significance thresholds set at $p = 0.05$. For the free recall task, inter-judge reliability was assessed using Cohen's kappa on 40% of the total dataset, yielding excellent reliability ($\kappa = 0.87$). The Aligned Rank Transform was used for inferential non-parametric analysis [12], followed by post-hoc tests using the Kenward-Roger method.

III. RESULTS OF EXPERIMENT 1

Velocity. The analysis revealed a main session effect ($F(2, 1492) = 19.72$, $p < 0.001$): post-hoc tests indicated that patterns were traced faster in the pre-test than in the short-term ($p < 0.001$) and long-term ($p < 0.001$) post-tests.

Trace length. The analysis revealed a main session effect ($F(2, 1492) = 33.25$, $p < 0.001$): post-hoc tests indicated that the trace of patterns was increased more in the pre-test than in the short-term ($p < 0.001$) and long-term ($p < 0.001$) post-tests, and more in the long-term post-test than in the short-term post-test ($p < 0.001$).

Movement dysfluency. No significant effects were revealed.

Spatial error. The analysis revealed a main session effect ($F(2, 1487) = 7.58$, $p < 0.001$): post-hoc tests indicated that spatial error was higher in the pre-test than in the short-term ($p < 0.05$) and long-term ($p < 0.001$) post-tests.

Number of correctly recalled graphic patterns. The analysis revealed a main effect of visual strategy ($F(1, 93) = 3.98$, $p = 0.04$): patterns learned without a written trace were better recalled than those learned with a written trace.

Number of correctly recognized graphic patterns. No significant effects were revealed.

IV. INTERMEDIATE DISCUSSION

The results indicated that training led the participants to reduce their velocity in favor of better spatial precision. These results are consistent with the well-known speed-accuracy trade-off. Furthermore, the results indicated a positive effect of the visual strategy on the free recall post-test, which is consistent with the literature [3]. One hypothesis is that a visual representation of the patterns is maintained in working memory to trace them correctly, requiring more attention and working memory resources, which enhances their transfer to long-term memory [13]. This visual strategy was therefore adopted for the next experiment.

V. Method of Experiment 2

A. Participants

Another group of 32 adult participants took part in this experiment (mean age: 24.84 ± 6.29, 25 female, and 31 right-handed), with the same inclusion criteria and informed consent form as in Experiment 1.

B. Task and procedure

The task and the material were largely the same as in the first experiment. The procedure was as follows (Fig. 3):

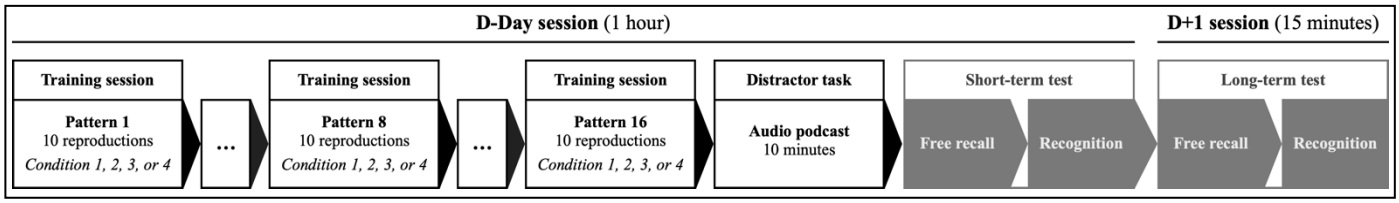


Fig 3. Experimental design of Experiment 2.

Training session. After the familiarization phase, the 16 graphic patterns were presented successively as in Experiment 1. Participants had to trace each one ten times. These patterns were evenly distributed across four writing conditions according to the presence or absence of auditory and haptic strategies (Table II):

- Audiohaptic: with index finger and with movement sonification.
- Auditory: with stylus and with movement sonification.
- Haptic: with index finger and without sonification.
- Control: with stylus and without sonification (i.e., without sensory strategy).

The conditions and presentation order were counterbalanced as in Experiment 1.

TABLE II. Writing conditions of Experiment 2.

Condition	Sound	Haptic	Strategie
1	Yes	Index finger	**Audiohaptic**
2	Yes	Stylus	**Auditory**
3	No	Index finger	**Haptic**
4	No	Stylus	**Control**

In all conditions, the written trace was invisible during reproduction and displayed for 3 seconds after completion. The sonification strategy was the same as in the first experiment.

Distractor task. Participants had to listen to a 10-minute audio podcast to minimize the recency effect.

Short-term and long-term tests. First, a free recall task was presented. Participants had to trace from memory as many patterns as they had learned during the training session. Then, a recognition task, which consisted in a lexical decision task, was presented. The 16 graphic patterns and their 32 distractors were randomly displayed, one at a time, for 3 seconds each. Participants had to determine as quickly as possible whether each pattern corresponded to the one they had learned or not.

C. Data analysis

Only one aspect was evaluated here: memorization. It was evaluated by calculating the number of graphic patterns correctly recalled/recognized divided by the number of patterns learned in each condition (as a ratio between 0 and 1), and the mean response time (in ms) for correctly recognized graphic patterns. The same statistical analyses as in Experiment 1 were carried out.

VI. Results of Experiment 2

Number of correctly recalled graphic patterns. The analysis revealed a main effect of auditory strategy ($F(1, 217) = 4.44$, $p = 0.03$): patterns learned with movement sonification were less well recalled than those learned without sonification (Fig. 4A).

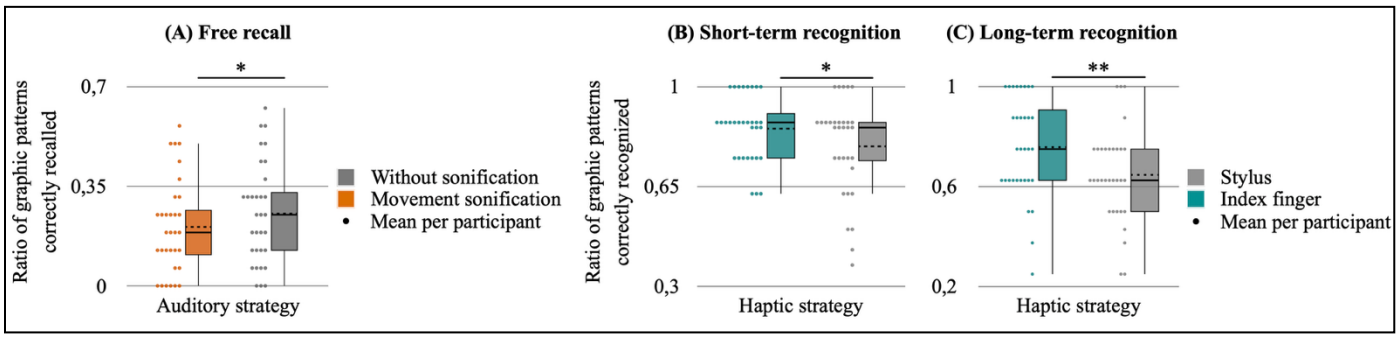


Fig 4. Main performance of Experiment 2.

Number of correctly recognized graphic patterns. The analysis revealed a main session effect ($F(1, 973) = 61.96$, $p < 0.001$): patterns were better recognized in the short-term test than in the long-term test. A session-by-session analysis revealed a main effect of haptic strategy in both the short-term ($F(1, 466) = 4.06$, $p = 0.04$) and long-term ($F(1, 476) = 8.53$, $p < 0.01$) tests: patterns learned with the index finger were better recognized than those learned with the stylus (Fig. 4B; Fig. 4C).

Response time of correctly recognized graphic patterns. No significant effects were revealed.

VII. Intermediate discussion

On the one hand, the results indicated a positive effect of the haptic strategy on the recognition task at both short-term and long-term tests, which is consistent with the literature [5]. One hypothesis could be that the finger enhances proprioceptive feedback information compared to the stylus, since this information is more solicited in the absence of immediate visual feedback. On the other hand, the results indicated a negative effect of the auditory strategy on the free recall. One hypothesis could be that the sounds associated with the spatial evolution of handwriting movement were not sufficiently discriminable from each other. This is consistent with the phonological similarity effect, which postulates that words associated with similar sounds are less well learned than those associated with distant sounds [14].

VIII. General discussion

Overall, no multisensory benefits were revealed from both experiments. Three main hypotheses can be considered. First, the multisensory integration of information that leads the writer to focus on the writing process rather than the writing product was probably not relevant given the nature of the learning task, which involved memorizing a spatial pattern. Second, the sonification strategy may not have been optimal. The primary benefit of sonification appears to be in enhancing movement dynamics rather than spatial memorization, considering that the auditory modality is less precise for spatial information [15]. Based on the Bayesian approach to multisensory integration, it is plausible that integration occurred, but that multisensory information was more imprecise than unisensory information. Furthermore, spatial sonification may not have been the most relevant strategy for improving the handwriting movement, in contrast to the beneficial effects of sonification of movement velocity revealed in previous studies ([16], for example). Finally, the multisensory condition may also have induced a problem of cognitive load. Based on the cognitive load theories [17], a redundancy effect occurs when identical information is presented simultaneously through multiple modalities, leading to unnecessary cognitive processing and potential overload. Similarly, a split-attention effect arises when learners must divide their attention between multiple sources of information, which can hinder learning. As a result, the sounds may have had a distracting effect. Therefore, careful design of multisensory integration is crucial to ensure they enhance learning without imposing excessive cognitive demands in the light of cognitive load theories. Exploring this multisensory strategy in clinical contexts—where visual feedback is absent or severely impaired, and other sensory modalities are used for acquiring spatial information—would be more relevant. From this perspective, investigating the effects of multisensory integration in Braille learning among visually impaired individuals could open up promising directions.

References

[1] Danna, J., & Velay, J.-L. (2015). Basic and supplementary sensory feedback in handwriting. *Frontiers in Psychology, 6*(169).

[2] Karlsdottir, R., & Stefansson, T. (2002). Problems in Developing Functional Handwriting. *Perceptual and Motor Skills, 94*(2), 623–662.

[3] Bara, F., & Bonneton-Botté, N. (2021). Handwriting isolated cursive letters in young children: Effect of the visual trace deletion. *Learning and Instruction, 74*(1), 101439.

[4] Danna, J., Paz-Villagrán, V., Capel, A., Pétroz, C., Gondre, C., Pinto, S., Thoret, E., Aramaki, M., Ystad, S., Kronland-Martinet, R., & Velay, J.-L. (2014). Movement Sonification for the Diagnosis and the Rehabilitation of Graphomotor Disorders. In M. Aramaki, O. Derrien, R. Kronland-Martinet, & S. Ystad (Eds.), *Sound, Music, and Motion* (Vol. 8905, pp. 246–255). Springer International Publishing.

[5] Patchan, M. M., & Puranik, C. S. (2016). Using tablet computers to teach preschool children to write letters: Exploring the impact of extrinsic and intrinsic feedback. *Computers & Education, 102*, 128–137.

[6] Versace, R., Vallet, G. T., Riou, B., Lesourd, M., Labeye, E., & Brunel, L. (2014). Act-In: An integrated view of memory mechanisms. *Journal of Cognitive Psychology, 26*(3), 280–306.

[7] Schneider, T. R., Engel, A. K., & Debener, S. (2008). Multisensory identification of natural objects in a two-way crossmodal priming paradigm. *Experimental psychology, 55*(2), 121–132.

[8] Seyll, L., Wyckmans, F., & Content, A. (2020). The impact of graphic motor programs and detailed visual analysis on letter-like shape recognition. *Cognition*.

[9] Effenberg, A. O., Schmitz, G., Baumann, F., Rosenhahn, B., & Kroeger, D. (2015). SoundScript – Supporting the Acquisition of Character Writing by Multisensory Integration. *The Open Psychology Journal*, *8*(1), 230-237.

[10] Danna, J., Paz-Villagrán, V., & Velay, J.-L. (2013). Signal-to-Noice velocity peaks difference: A new method for evaluating the handwriting movement fluency in children with dysgraphia. *Research in Developmental Disabilities, 34*(12), 4375–4384.

[11] Connan, J. F., Jover, M., Saint-Cast, A., & Danna, J. (2023). Does modifying visual feedback facilitate learning to write new pseudoletters?. *Human movement science*, *87*, 103046.

[12] Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The Aligned Rank Transform for nonparametric factorial analyses using only ANOVA procedures. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York*, 143–146.

[13] Connan, J.-F., Jover, M., Saint-Cast, A., & Danna, J. (2021). Comment les nouvelles technologies peuvent-elles aider le scripteur à mieux écrire ? Étude pilote sur la modification de la perception visuelle de la trace. *ANAE - Approche Neuropsychologique des Apprentissages chez l'Enfant, 170*, 77–89.

[14] Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *The Quarterly Journal of Experimental Psychology, 18*(4), 362–365.

[15] Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*(3), 638–667.

[16] Danna, J., Fontaine, M., Paz-Villagrán, V., Gondre, C., Thoret, E., Aramaki, M., Kronland-Martinet, R., Ystad, S., & Velay, J.-L. (2015). The effect of real-time auditory feedback on learning new characters. *Human Movement Science, 43*, 216–228.

[17] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285.

Influence of Sampling Rate on Handwriting in Children with Developmental Dysgraphia

Jan Mucha*‡, Jiri Mekyska*†, Katarina Zvoncakova†, Adam Klocek†
*Department of Telecommunications, Brno University of Technology, Brno, Czech Republic
Email: mucha@vut.cz
†Department of Research Methodology, Institute of Psychology, Czech Academy of Sciences, Brno, Czech Republic
‡Tecnocampus Universitat Pompeu Fabra, Mataró, Spain

***Abstract*— Digital handwriting analysis provides objective, quantitative measures of fine m otor s kills w ith g rowing a pplication i n the assessment of neurodevelopmental disorders. However, standardized data acquisition protocols remain underdeveloped. This study investigates the impact of sampling frequency on kinematic features of handwriting in children with Developmental Dysgraphia (DD; n = 155) and Intact Controls (IC; n = 114). Participants completed three graphomotor tasks and a paragraph copy task on a Wacom Cintiq 16 tablet. Original signals (fs = 187 Hz) were systematically downsampled from 170 Hz to 10 Hz. Ninety-four kinematic, dynamic, spatial, and temporal features were extracted. Mixed-design ANOVAs (sampling frequency × diagnosis) were performed on the Normalized Root Mean Squared Error (NRMSE) between original and downsampled signals for each feature. Significant i nteractions b etween s ampling f requency a nd diagnosis were observed for several kinematic features: overall and horizontal on-surface velocity and overall and vertical in-air velocity. These findings indicate that downsampling disproportionately affected the DD group.

Keywords*— *developmental dysgraphia, graphomotor disabilities, sampling rate, sampling frequency, online handwriting, down-sampling

I. Introduction

Digital tools and technologies are increasingly transforming the assessment and analysis of psychological and neurological states. This trend is particularly evident in studies of neurodevelopmental and neurodegenerative disorders, where objective, quantifiable measures are essential for early diagnosis, monitoring disease progression, and evaluating treatment efficacy [1], [2]. Online and remote assessment methods are growing, offering the potential for broader access to care and facilitating longitudinal studies [3]. Reflecting this broader trend, our research has focused on applying digital handwriting analysis to characterize and differentiate neurological conditions, including Parkinson's disease, Dementia with Lewy Bodies, and developmental disorders such as Developmental Dysgraphia (DD) in children [4], [5].

Handwriting, a complex fine motor skill involving the coordination of cognitive, perceptual, and motor processes, is often impaired in children with neurodevelopmental conditions such as Developmental Dysgraphia (DD). This disability is a specific learning disorder characterized by impaired handwriting fluency and legibility despite adequate cognitive abilities [6]. Affected children struggle with letter formation and broader graphomotor tasks [4]. These difficulties stem from impairments in motor control, visual-motor integration, and/or orthographic coding—the ability to represent the visual form of letters and words. Children with DD exhibit less automated writing and drawing movements, requiring greater effort and control than their typically developing peers [7], [8]. The kinematic handwriting characteristics, readily captured using digital devices, can offer insights into these underlying motor inefficiencies.

Traditionally, handwriting assessment relied on qualitative observations by clinicians. However, the emergence of digital tablets and specialized software has enabled precise measurement of handwriting characteristics, including kinematic, dynamic, spatial, or in-air movement features [9]. These parameters offer objective and quantifiable indicators of handwriting performance, facilitating a more nuanced and sensitive analysis of motor control deficits. Recent advancements, such as the DiaGraMo system [10], have focused on developing online tools for objective assessment of graphomotor difficulties. This trend towards remote, digital assessment holds significant promise for improving access to care, especially for individuals in underserved areas or with limited mobility.

Despite growing interest in digital handwriting analysis and the proliferation of online assessment tools, limited research has addressed data quality in acquired handwriting signals. Specifically, the impact of a digitizing tablet's sampling frequency on the fidelity of the captured signal and the subsequent effects on derived kinematic features remain underexplored. While higher sampling frequencies capture more detailed information, they also increase data storage and computational demands. Conversely, low sampling frequencies may lead to information loss, potentially obscuring the handwriting signal's subtle but clinically relevant features. We hypothesize that inadequate sampling frequencies can introduce variability and noise into handwriting data, thus impacting the validity and reliability of diagnostic and research findings. This inconsistency hinders the ability to detect meaningful differences between groups (e.g., children with DD versus intact children (IC)) and to track changes over time. The lack of standardized guidelines for handwriting data acquisition represents a significant gap in the field, potentially limiting the comparability and reproducibility of research findings [4].

Therefore, our primary motivation in this study is to advance the discussion within the scientific community regarding establishing digital handwriting signal acquisition standards. To address this, the primary aim of this paper is to preliminarily determine the impact of sampling frequency on the quality of handwriting features and to investigate potential differences between children with DD and IC. The findings will contribute to identifying a "cut-off" sampling frequency that balances data fidelity with practical constraints, thereby promoting more robust and comparable research in digital handwriting analysis.

II. Materials and Methods

A. Database

In this study, we used a dataset comprising handwriting samples from children in Czech primary schools. A total of 269 children participated, stratified into two groups: IC and children diagnosed with DD by a special education needs teacher. The IC group included 50 third-grade students (age 9.1 $\pm$ 0.5 years; 34 boys, 16 girls) and 64 fourth-grade students (age 10.2 $\pm$ 0.5 years; 50 boys, 14 girls). The DD group consisted of 59 third-grade students (age 9.1 $\pm$ 0.5 years; 38 boys, 21 girls) and 96 fourth-grade students (age 10.2 $\pm$ 0.5 years; 75 boys, 21 girls).

Data was acquired using a Wacom Cintiq 16 (DTK1660K0B) display tablet. Participants were instructed to complete four graphomotor tasks (upper loops fast, combined loops precise, lower loops precise) and one writing task (a paragraph copy task, with different versions for third and fourth graders). These tasks were chosen to elicit handwriting movements that allow for the detailed analysis of kinematic parameters, such as velocity, acceleration, and fluency, across varying levels of task complexity.

The parents of all participating children provided informed consent, and the Ethics Committee of the Institute of Psychology at the Czech Academy of Sciences approved the study. All procedures adhered to the Ethical Principles of Psychologists and Code of Conduct outlined by the American Psychological Association.

B. Feature Extraction

The handwriting data were processed to extract a comprehensive set of kinematic, dynamic, spatial, and temporal features. Where applicable, these features were computed separately for both on-surface (pen-down) and in-air (pen-up) movements. All features were calculated using custom Python scripts, leveraging the *handwriting-sample* and *handwriting-features* packages.

The original sampling frequencies of the handwriting samples varied across participants, with a mean of 187 Hz. To investigate the impact of sampling frequency on feature quality, we systematically downsampled the original signals to target frequencies of 170, 150, 130, 110, 90, 70, 50, 30, and 10 Hz. Downsampling was performed using linear interpolation to preserve the overall shape of the signals.

We computed a range of features for each handwriting sample (both original and downsampled). Kinematic features included velocity (overall, horizontal, and vertical components), acceleration (overall, horizontal, and vertical), jerk (overall, horizontal, and vertical), and the relative number of changes in the velocity profile as a measure of movement smoothness. Dynamic features comprised pen pressure, azimuth, and tilt angles. Spatial features included stroke length, height, and width, calculated for both on-surface and in-air strokes where applicable. Temporal features encompassed stroke duration (on-surface and in-air) and the relative number of interruptions. The mean, standard deviation, and median were computed for all vector-based features, yielding scalar values for each participant, task, and sampling frequency. The relative number of changes in the velocity profile and the relative number of interruptions were calculated as single scalar values. This process resulted in 94 handwriting features for each sampling frequency.

Following an initial correlation analysis (described in the next section), we focused on quantifying the impact of downsampling on these features. To this end, we computed several metrics comparing the original features to the corresponding features at each downsampled frequency: root mean squared error (RMSE), normalized RMSE (NRMSE), and the mean and standard deviation of the absolute difference. These metrics were calculated for each feature, participant, and downsampled frequency. RMSE and absolute differences provide measures of the overall magnitude of the discrepancy between the original and downsampled signals. The NRMSE was calculated as follows:

$$\text{NRMSE} = \frac{\text{RMSE}}{\max(X_{original}) - \min(X_{original})} \tag{1}$$

where $X_{original}$ represents the time series of the original feature, normalizes the RMSE by the range of the original signal. This normalization provides a scale-invariant measure of the error, allowing for comparisons across features with different units and magnitudes.

C. Statistical Analysis and Machine Learning

Prior to statistical analysis, we considered grade and gender as confounding factors and removed their effects using linear regression [11]. Subsequently, we calculated Spearman's correlation (with False Discovery Rate (FDR) correction for multiple

comparisons) to explore the relationships between handwriting features, clinical group (IC/DD), and, for the paragraph copy task, handwriting quality as measured by a Handwriting Legibility Scale (HLS) [12].

To evaluate the impact of sampling frequency on handwriting features and to compare performance between the DD and IC groups, we employed a mixed-design Analysis of Variance (ANOVA). This approach was chosen because it allows for the simultaneous analysis of both within-subjects factors (sampling frequency) and between-subjects factors (diagnosis group). The dependent variable in all ANOVA models was the NRMSE.

Prior to conducting the ANOVA, the normality of the data was assessed using the Shapiro-Wilk test ($p < 0.05$). Mauchly's test of sphericity was performed for each ANOVA to assess the assumption of sphericity for the within-subjects factor. For the interaction effect, the uncorrected p-value (p-unc) was used. For any feature exhibiting a significant interaction effect ($p < 0.05$), post-hoc pairwise comparisons were performed with Bonferroni correction for multiple comparisons.

III. Results

Table I presents Spearman's correlation (ρ) and FDR-corrected p-values for selected handwriting features across all sampling frequencies, correlated with either the clinical group (DD/IC) or the HLS score (for copy tasks only). For the upper loops fast task, mean pressure showed significant correlations with the clinical group at all frequencies. For the combined loops precise task, the standard deviation of on-surface tilt showed significant correlations with the clinical group at most frequencies, except 170 Hz. The mean on-surface stroke duration for the lower loops precise task showed significant negative correlations with the clinical group at all frequencies. For the third-grade copy task, mean in-air velocity correlated with the HLS close to the significant level at the original frequency, with significance maintained down to 130 Hz. For the fourth-grade copy task, the relative number of velocity changes showed a significant negative correlation with the HLS score at the original frequency, with significance maintained down to 110 Hz.

Table II presents the significant interaction effects (sampling frequency $\times$ diagnosis IC/DD) found in the mixed-design ANOVAs. The table includes the F-statistic (F), numerator (df1) and denominator (df2) degrees of freedom, uncorrected p-value (p-unc), partial eta-squared (η_p^2), and the sampling frequencies showing significant group differences in post-hoc tests.

Five features showed significant interactions. For the lower loops precise task, a significant interaction was found for horizontal on-surface velocity ($F_{(8,1744)} = 4.86$, $p < 0.001$, $\eta_p^2 = 0.022$). Post-hoc comparisons revealed significant differences between the DD and IC groups at 10 Hz, 30 Hz, 50 Hz, and 70 Hz. For the fourth-grade copy task, significant interactions were found for both overall on-surface velocity ($F_{(8,112)} = 3.21$, $p = 0.003$, $\eta_p^2 = 0.187$) and horizontal on-surface velocity ($F_{(8,112)} = 3.21$, $p = 0.003$, $\eta_p^2 = 0.187$). Post-hoc tests for on-surface velocity revealed significant group differences at 30, 50, 90, and 150 Hz, while for horizontal on-surface velocity, significant differences were found at 30, 50, 90, 110, and 150 Hz. For the upper loops fast task, significant interactions were found for both overall in-air velocity ($F_{(8,136)} = 2.18$, $p = 0.033$, $\eta_p^2 = 0.114$) and vertical in-air velocity ($F_{(8,136)} = 2.26$, $p = 0.027$, $\eta_p^2 = 0.117$); for both features post-hoc test showed significant group difference at 10 Hz.

TABLE I. Selected Results per Task of the Spearman's Correlation Analysis

Task	Feature	Scale	original		170 Hz		150 Hz		130 Hz		110 Hz		90 Hz		70 Hz		50 Hz		30 Hz		10 Hz	
			ρ	adj p	ρ	adj p	ρ	adj p	ρ	adj p	ρ	adj p	ρ	adj p	ρ	adj p	ρ	adj p	ρ	adj p	ρ	adj p
lp_{up}	press (m)	IC/DD	0.22	0.008	0.21	0.027	0.21	0.027	0.23	0.007	0.23	0.003	0.22	0.007	0.22	0.007	0.22	0.010	0.22	0.007	0.22	0.007
lp_{com}	$tilt_s$ (s)	IC/DD	0.23	0.010	0.19	0.100	0.22	0.025	0.22	0.017	0.22	0.010	0.23	0.010	0.23	0.010	0.23	0.010	0.23	0.010	0.23	0.010
lp_{low}	$strk_d_s$ (m)	IC/DD	-0.19	0.013	-0.18	0.062	-0.16	0.076	-0.18	0.030	-0.19	0.016	-0.19	0.014	-0.19	0.012	-0.19	0.008	-0.19	0.009	-0.19	0.010
cp_3	vel_a (m)	HLS	0.23	0.042	0.00	1.000	0.14	0.777	0.25	0.099	0.29	0.007	0.26	0.013	0.25	0.019	0.23	0.038	0.22	0.043	0.21	0.052
cp_4	rel_num_vel	HLS	-0.31	0.001	-0.14	0.988	-0.24	0.533	-0.28	0.009	-0.22	0.019	-0.18	0.073	-0.14	0.186	-0.08	0.509	–	–	–	–

* ρ – Spearman's correlation coefficient; adj p – adjusted p-value of Spearman's correlation; IC/DD – diagnosis state; lp_{up} – upper loops fast; lp_{com}– combined loops fast; lp_{low} – lower loops precise; cp_3 – copy task for third grade; cp_4– copy task for fourth grade; press – pressure; $tilt_s$ – tilt on surface; $strk_d_s$ – stroke duration on surface; vel_a – velocity in air; rel_num_vel – relative numbers of changes in velocity profile; (m) – mean; (s) – std.

IV. Discussion and Conclusion

Initial correlation analyses revealed generally weak, though sometimes statistically significant, associations between handwriting features and clinical group or handwriting quality. While average Spearman's ρ values were below 0.3, spatial, dynamic, and temporal features exhibited relatively high resistance to downsampling. This is attributed to their representation of more global aspects of handwriting production, which are less sensitive to the fine-grained temporal variations captured by higher frequencies.

Given the more significant impact of downsampling on kinematic features observed in the correlation analysis, subsequent ANOVA analyses focused on velocity, acceleration, and jerk. Significant interactions ($p < 0.05$) between sampling frequency

TABLE II. SUMMARY OF SIGNIFICANT INTERACTION EFFECTS FROM MIXED-DESIGN ANOVA

Feature	Task	Source	F	df1	df2	p-unc	η_p^2	Significant Frequencies
velocity horizontal on surface	lower loops precise	Interaction	4.86	8	1744	0.001	0.022	10, 30, 50, 70
velocity on surface	copy task for fourth grade	Interaction	3.21	8	112	0.003	0.187	30, 50, 90, 150
velocity horizontal on surface	copy task for fourth grade	Interaction	3.21	8	112	0.003	0.187	30, 50, 90, 110, 150
velocity vertical in air	upper loops fast	Interaction	2.26	8	136	0.027	0.117	10
velocity in air	upper loops fast	Interaction	2.18	8	136	0.033	0.114	10

* Only features with a significant interaction effect (sampling frequency × diagnosis, p ¡ 0.05) are included. df1 and df2 represent the numerator and denominator degrees of freedom, respectively. η_p^2 represents partial eta-squared. Significant Frequencies shows the list of frequencies after post-hoc analysis with significant differences.

and diagnosis were found for several kinematic features: horizontal on-surface velocity ($F_{(8,1744)}$ = 4.86); overall and horizontal on-surface velocity ($F_{(8,112)}$ = 3.21); and overall and vertical in-air velocity ($F_{(8,136)}$ = 2.18 and $F_{(8,136)}$ = 2.26, respectively). These interactions indicate a differential effect of reduced sampling frequency on these measures between children with DD and IC. The DD group exhibited significantly more significant performance degradation (measured by NRMSE) at lower sampling frequencies. For example, the mean NRMSE for horizontal on-surface velocity in the lower loops precise task increased by 10% for the DD group between 180 Hz and 30 Hz, compared to a 7% increase for the IC group. This suggests that higher-frequency information is critical for characterizing the handwriting of children with DD, likely because DD involves less automated and more variable movements [4], [13], which are obscured at lower sampling rates.

This study's limitations include its preliminary nature, the use of linear interpolation for downsampling (which differs from straightforward data point dropping), and the limited set of tasks (n=4) and features (n=94) examined. Future research should include classification analysis and investigate the impact of sampling frequencies in a broader range of handwriting tasks and features. Nevertheless, this work represents a step toward establishing digital handwriting data acquisition standards. The findings highlight the sensitivity of kinematic features to sampling frequency, particularly when assessing children with DD. Further research should confirm and refine these preliminary results.

ACKNOWLEDGMENT

This work was supported by the Czech Science Foundation, project no. GN23-06074O (Research of advanced methods of graphomotor disabilities analysis based on fractional calculus), by the project Mediated Society (MEDIS:ON) CZ.02.01.01/00/23_025/0008713 co-financed/supported by the European Union and Czech Republic, and projects no. 25-15668S supported by the Czech Science Foundation.

REFERENCES

[1] M. Choo, D. Park, M. Cho, S. Bae, J. Kim, and D. H. Han, "Exploring a multimodal approach for utilizing digital biomarkers for childhood mental health screening," *Frontiers in psychiatry*, vol. 15, p. 1348319, 2024.

[2] H. Motahari-Nezhad, M. Fgaier, M. Mahdi Abid, M. Péntek, L. Gulácsi, and Z. Zrubka, "Digital biomarker–based studies: scoping review of systematic reviews," *JMIR mHealth and uHealth*, vol. 10, no. 10, p. e35722, 2022.

[3] J. M. Hatcher-Martin, N. A. Busis, B. H. Cohen, R. A. Wolf, E. C. Jones, E. R. Anderson, J. V. Fritz, S. J. Shook, and R. M. Bove, "American academy of neurology telehealth position statement," *Neurology*, vol. 97, no. 7, pp. 334–339, 2021.

[4] J. Mekyska, K. Safarova, T. Urbanek, J. Bednarova, V. Zvoncak, J. M. Havigerova *et al.*, "Graphomotor and handwriting disabilities rating scale (GHDRS): Towards complex and objective assessment," *Australian Journal of Learning Difficulties*, vol. 29, no. 1, pp. 1–34, 2024.

[5] J. Mucha, M. Gavenciak, J. Mekyska, M. Faundez-Zanuy, L. Brabenec, and I. Rektorova, "Evaluation of fractional calculus and delta parameters in prodromal diagnosis of dementia with Lewy bodies utilizing online handwriting," in *2024 32nd European Signal Processing Conference (EUSIPCO)*, 2024, pp. 1731–1735.

[6] P. J. Chung, D. R. Patel, and I. Nizami, "Disorder of written expression and dysgraphia: definition, diagnosis, and management," *Translational pediatrics*, vol. 9, no. Suppl 1, p. S46, 2020.

[7] T. Gargot, T. Asselborn, H. Pellerin, I. Zammouri, S. M. Anzalone, L. Casteran, W. Johal, P. Dillenbourg, D. Cohen, and C. Jolly, "Acquisition of handwriting in children with and without dysgraphia: A computational approach," *PloS one*, vol. 15, no. 9, p. e0237575, 2020.

[8] S.-H. Chang and N.-Y. Yu, "Handwriting movement analyses comparing first and second graders with normal or dysgraphic characteristics," *Research in developmental disabilities*, vol. 34, no. 9, pp. 2433–2441, 2013.

[9] J. Mekyska, Z. Galaz, K. Safarova, V. Zvoncak, J. Mucha, Z. Smekal, A. Ondrackova, T. Urbanek, J. M. Havigerová, J. Bednarova *et al.*, "Computerised assessment of graphomotor difficulties in a cohort of school-aged children," in *2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. IEEE, 2019, pp. 1–6.

[10] "DiaGraMo: Diagnostic System for Graphomotor Difficulties," https://shop.propsyco.cz/diagramo.

[11] M. T. Todd, L. E. Nystrom, and J. D. Cohen, "Confounds in multivariate pattern analysis: theory and rule representation case study," *Neuroimage*, vol. 77, pp. 157–165, 2013.

[12] E. Hamstra-Bletz, J. de Bie, and B. den Brinker, *Beknopte beoordelingsmethode voor kinderhandschriften*. Lisse: Swets & Zeitlinger, 1987, [Concise assessment method for children's handwriting].

[13] L. Deschamps, L. Devillaine, C. Gaffet, R. Lambert, S. Aloui, J. Boutet, V. Brault, E. Labyt, and C. Jolly, "Development of a pre-diagnosis tool based on machine learning algorithms on the BHK test to improve the diagnosis of dysgraphia," *Advances in Artificial Intelligence and Machine Learning*, vol. 1, no. 2, pp. 114–135, 2021.

Handwritten Character Representation and Generation Based on Bézier Curves

Aleksandra Dundic[1], Andreas Fischer[1-2], Anna Scius-Bertrand[2]

[1]*University of Fribourg, Bd de Pérolles 90, 1700 Fribourg, Switzerland*
[2]*University of Applied Sciences and Arts Western Switzerland, HEIA, Bd de Pérolles 80, 1700 Fribourg, Switzerland*

aleksandra.dundic@unifr.ch, andreas.fischer@unifr.ch, anna.scius-bertrand@hefr.ch

Abstract — Handwritten character recognition is often hindered by the limited availability of annotated training data for neural networks. While synthetic data has shown promise in addressing this issue, generating realistic handwriting variations remains challenging due to the high variability in handwriting styles. This article introduces a novel method for generating synthetic character variations by modeling skeletons and contours with cubic Bézier curves. By applying weighted Gaussian noise to the control points of these curves, we generate natural variations while preserving the structure of the characters. Contour-based variations are rendered by filling enclosed regions with black pixels, whereas skeleton-based variations retain essential geometric features. The effectiveness of the proposed approach is evaluated for neural network training, demonstrating its potential to improve handwritten character recognition performance.

Keywords — Handwritten character recognition, Bézier curves, synthetic data generation, Hausdorff distance

I. Introduction

Handwritten character recognition plays a key role in preserving cultural heritage, yet achieving high accuracy remains challenging due to handwriting variability. A major obstacle is the limited availability of annotated training data, which is essential for neural network training. Since manual data collection is time-consuming, data augmentation provides a practical solution for generating diverse handwriting variations. Data augmentation methods fall into traditional and generative approaches. Traditional techniques operate at the pixel level and include affine transformations (e.g., rotation, scaling, shearing), elastic distortions, and noise injection. These methods enhance model generalization by introducing controlled variations [1,2,3]. Generative approaches, such as Generative Adversarial Networks (GANs), synthesize handwriting samples by learning from existing datasets, demonstrating effectiveness in producing diverse styles [4].

Most augmentation techniques, particularly pixel-based ones, modify images without preserving structural integrity, such as smooth contours and consistent strokes. In contrast, curve-based character representation models letters as mathematical curves, enabling precise shape variations. For example, weighted quadratic Bézier curves have been proposed to reconstruct character strokes, capturing handwriting fluidity [5]. However, this method focuses on reconstruction rather than augmentation, and quadratic Bézier curves offer limited flexibility for complex shapes.

This paper introduces a novel approach using cubic Bézier curves to represent character skeletons and contours, providing greater control and accuracy for intricate shapes. Weighted Gaussian noise is applied to control points, generating variations of the Bézier curves that are then converted into synthetic character images for neural network training.

Section II details the method, Section III presents the experimental setup and results, and the final section discusses conclusions and future work.

II. Method

A schematic overview of the proposed method is shown in Figure 1. The initial letter prototype can be handwritten or synthetic, for example a printed font, and there are two possible letter representations:

- Skeleton-based: Letter representation based on medial axes with a width of one pixel.
- Contour-based: The area between inner and outer contour is filled black at the end of the process.

We consider the `skeletonize` and `findContours` methods from `OpenCV` to extract the skeleton and the contours, respectively, and remove branches of the skeleton if they are too small (threshold optimized experimentally). Afterwards, our method consists of two main steps, Bézier representation and making letter variations, as detailed below.

IGS 2025 ©2024 IEEE

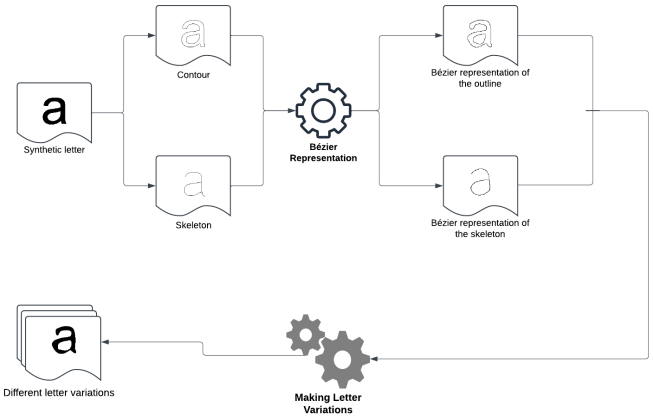


Figure 1. An overview of the proposed method.

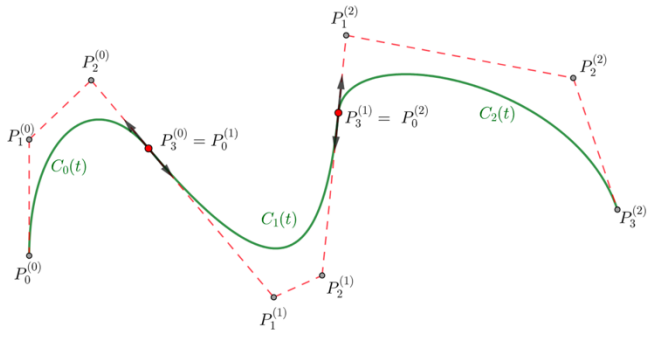


Figure 2. Piecewise cubic Bézier curve.

A. Bézier Representation

In order to obtain an analytical, modifyable representation of the letter, we individually process each segment of the skeleton between intersections and endpoints. For contour-based representation, the inner and outer contours are processed separately. The pixel chains are approximated with cubic Bézier curves using Schneider's algorithm [6]. It considers piecewise cubic Bézier curves as illustrated in Figure 2. In this example, the curve consists of three Bézier curves $C_0(t)$, $C_1(t)$, and $C_2(t)$, each of which has a starting point $P_0^{(i)}$, two control points $P_1^{(i)}$ and $P_2^{(i)}$, and an end point $P_3^{(i)}$. They are smoothly connected over end points that are at the same time starting points of the next Bézier curve, for example $P_3^{(0)} = P_0^{(1)}$. The algorithm computes the maximum error between the pixel chain and the current solution to determine if a split is needed for inserting an additional Bézier curve. The process is repeated until the maximum error falls below a predefined threshold, aiming at a balance between model simplicity (few Bézier curves) and model accuracy (low maximum error).

B. Letter Variations

Once letter skeletons or contours are represented using cubic Bézier curves, character variations are generated by modifying their control points. For ensuring realistic and visually coherent variations, three key constraints are applied: weighted Gaussian noise to preserve the general shape, Hausdorff distance to maintain structural integrity, and inner-outer contour resolution to prevent intersections. Different variation results are illustrated in Figure 4.

1) Weighted Gaussian noise

To introduce variability in letter shapes, Gaussian noise is added to each Bézier control point, as illustrated in Figure 3. However, uniform noise can cause distortions, as control points affect the curve differently. To address this, a weighting scheme based on the Euclidean distance (*dist*) between a control point (P) and its Bézier curve (C) is introduced:

$$weight(P) = \log(dist(P, C) + \varepsilon),\ \varepsilon \geq 1$$

That is, points close to the curve (close to zero distance), which greatly affect the shape when moved, receive almost no noise.

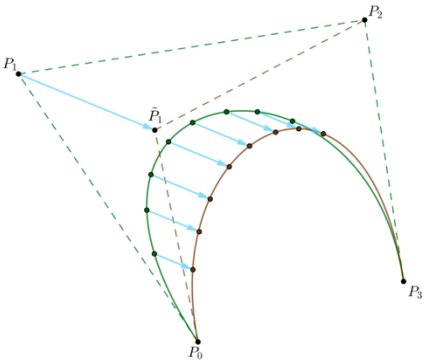


Figure 3. Moving a control point.

2) Hausdorff distance

While weighted noise improves deformation control, it does not fully prevent distortions like self-intersections. To address this, the Hausdorff distance between each generated variation B and its original Bézier representation A is computed:

$$d_H(A, B) = \max\{\max_{a \in A} \min_{b \in B} d(a, b), \max_{b \in B} \min_{a \in A} d(b, a)\}$$

Variations exceeding a predefined threshold are discarded.

3) Inner-outer contour intersection

While the previous steps are effective for skeletons, they are insufficient for letter contours with both inner and outer contours. Small perturbations can cause intersections when contours are too close. To resolve this, the outer contour is preserved between intersection points, while the corresponding inner contour portion is discarded. The space between the adjusted contours is then filled with black pixels, generating letter variations that mimic different writing styles or ink effects.

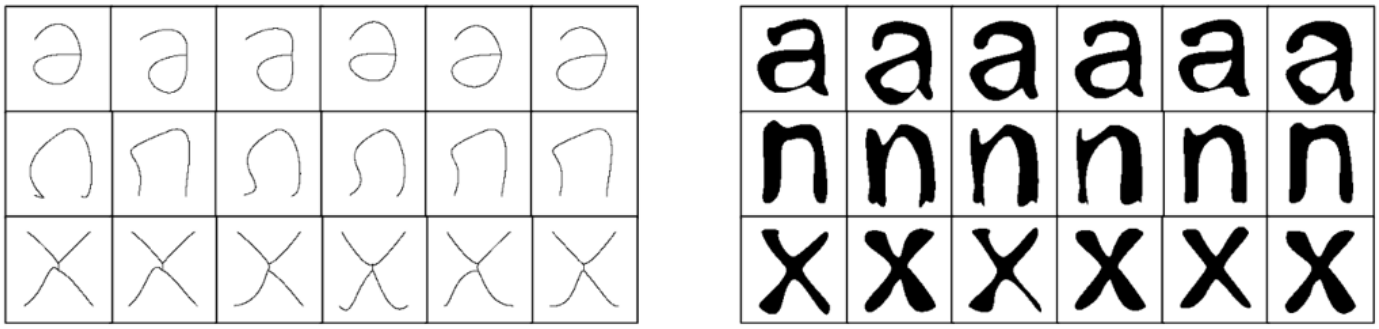


Figure 4. Variations of character skeletons (left) and contours (right) of letters 'a', 'n', and 'x'.

III. Experiements

To assess the impact of generating synthetic training sets using Bézier representations, we compare the handwritten recognition accuracy across different training strategies outlined in the experimental setup. After parameter optimization, the results are presented below.

A. Experimental setup

Experiments were conducted on a subset of the EMNIST dataset, which contains handwriting images of letters and digits. We limit the experiments to 26 lowercase letters and resize each image to 128 × 128 pixels for our algorithm. The goal is to classify the handwriting images using shallow convolutional neural networks (CNN).

1) Training, validation, and test sets

We distinguish two training scenarios:

- Train-10: using 10 training samples per letter class
- Train-100: using 100 training samples per letter class

The validation set for hyper-parameter optimization contains 10 images per letter (260 images in total) and the test set 100 images per letter (2600 images in total). The training, validation, and test sets are distinct and do not share any handwriting image.

2) Synthetic data generation

To assess the impact of generating synthetic data with our proposed method, we consider 26 letter prototypes from the Arial font. They are used to generate five synthetic training sets (T, 2T, 3T, 4T, and 5T) for each training scenario with the same size as the training set (T), double the size (2T), etc. They are added to the original handwriting images.

3) Standard data augmentation

To compare the effect of the proposed synthetic data generation with standard data augmentation techniques, we include rotation, shear, and Gaussian blur during training.

B. Hyper-parameter optimization

The CNN architecture is optimized over 3, 4, and 5 layers of convolutions, each followed by 2x2 max pooling. A growing number of 3x3 filters are considered, i.e. [16, 32, 64], [16, 32, 64, 128], and [16, 32, 64, 128, 256], followed by a dense layer with 50% dropout. CNNs are trained over 10 epochs with cross-entropy loss, Adam optimizer, learning rate 0.001, and batch size 32.

For the synthetic data generation, the size of the synthetic dataset is optimized over {T, 2T, 3T, 4T, 5T}, the standard deviation of the Gaussian noise over {10, 15, 20, 30, 40, 50} and the Hausdorff threshold over {10, 15, 20, 25, 30, 40}.

Regarding standard data augmentation, the maximum rotation range is optimized over {0, 5, 10, 15} degrees, the sheer range over {0, 0.1, 0.2}, and the Gaussian blur over {None, 3, 5} kernel size.

The optimal parameters are determined with respect to the classification accuracy on the validation set. They are listed in Table I.

TABLE I. Hyper-Parameters

Hyper-Parameters	Skeletons	Contours
CNN: filter configuration	[16, 32, 64, 128]	[16, 32, 64, 128, 256]
Synthetic data generation: data size, standard deviation, Hausdorff threshold	4T, 20, 10	2T, 20, 10
Standard data augmentation: rotation, shearing, Gaussian blur	10, 0, None	15, 0.2, 5

C. Results

Tables II summarizes the character classification accuracy results on the test set for skeletons and contours, respectively. For both training scenarios Train-10 and Train-100, three experiments are conducted: using only the original handwriting images, adding synthetic data using our proposed method, and additionally performing standard data augmentation.

TABLE II. TEST ACCURACY (PERCENTAGE) FOR SKELETONS AND CONTOURS.

Letter representation	Training scenario	Original	Original + Synthetic	Original + Synthetic + Augmentation
Skeletons	*Train-10*	58.1	73.3	73.7
	Train-100	93.0	97.4	95.4
Contours	*Train-10*	54.2	70.8	74.2
	Train-100	85.2	87.6	88.8

The proposed method using Bézier curve-based synthetic variations substantially enhances performance. Accuracy improves to 70.8% (Train-10) and 87.6% (Train-100) for contour-based models. Skeleton-based models achieve even better results, with 73.3% in Train-10 and a peak of 97.4% in Train-100, demonstrating our method's effectiveness.

Combining synthetic variations with standard augmentation generally improves the performance further. The best accuracy for contour models reaches 74.2% (Train-10) and 88.8% (Train-100). For skeleton-based models, the best result is 73.7% (Train-10), but a slight decrease is observed for Train-100 (95.4%), highlighting the need for careful tuning of augmentation parameters when combined with synthetic variations.

IV. CONCLUSION

This study investigates the application of cubic Bézier curves for handwriting recognition and generation, modeling skeletons or outlines as curve sequences to enrich training data. On a 26 characters subset of the EMNIST benchmark, we have observed significant improvements of the character recognition accuracy when adding handwriting-like variations of synthetic letter prototypes to the original handwriting training samples. When training data is limited, i.e. only 10 learning samples per letter class, the observed improvement is particularly strong.

The present paper opens up several lines of future research. To generalize our approach, we aim to apply the synthetization to entire words and include a wider range of datasets for handwriting recognition. Furthermore, the initial letter prototypes used in this study are synthetic, i.e. letters of a printed font. It would be interesting to investigate real handwriting for the letter prototypes, allowing the recognition model to adapt to specific handwriting styles. Finally, in order to achieve more natural variations, it would be promising to adapt the random noise on the control points towards real handwriting variability, e.g. identifying stable and instable control points on real data.

V. BIBLIOGRAPHY

[1] P. J. Simard, D. Steinkraus, and J. C. Platt, *Best practices for convolutional neural networks applied to visual document analysis*, in Proc. International Conference on Document Analysis and Recognition (ICDAR), 2003, pp. 958–962.

[2] O. Golubitsky, V. Mazalov, and S. M. Watt, *Toward affine recognition of handwritten mathematical characters*, in Proc. International Workshop on Document Analysis Systems, June 2010, pp. 35–42.

[3] Y. Bengio, F. Bastien, A. Bergeron, A. Boulanger-Lewandowski, Y. Chherawala, M. Cissé, M. Côté, D. Erhan, J. Eustache, X. Glorot, *Deep self-taught learning for handwritten character recognition*, in NIPS 2010 Deep Learning and Unsupervised Feature Learning Workshop, September 2010.

[4] E. Alonso, B. Moysset, and R. Messina, *Adversarial Generation of Handwritten Text Images Conditioned on Sequences*, in arXiv preprint arXiv:1903.00277, 2019.

[5] Z. Huang, W. Heng, Y. Tao, and S. Zhou, *Stroke-based character reconstruction*, in arXiv preprint arXiv:1806.08990, 2018.

[6] P. J. Schneider, *An algorithm for automatically fitting digitized curves*, in Graphics Gems, Academic Press, pp. 612–626, 1990.

Write on Paper and Get the Online Digital Trace: A New Era for Handwriting

Florent Imbert[1,5], Yann Soullard[2], Eric Anquetil[1], Tanja Harbaum[3], Alexey Serdyuk[3], Fabian Kreß[3], Tim Hamann[4], Peter Kämpf[4]
[1]IRISA, Université de Rennes, INSA Rennes, France
[2]IRISA, Université Rennes 2, France
[3]Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
[4]STABILO International GmbH, Heroldsberg, Germany
[5]LTU, Machine Learning Group, Luleå University of Technology, Sweden
yann.soullard@irisa.fr

***Abstract*— Capturing the digital trace of handwriting usually requires a specific stylus and a compatible substrate, be it acapacitive touchscreen, an ElectroMagnetic Resonance (EMR) tablet as used in Wacom systems or special paper. While writing on regular paper offers rich haptics, no latency and is well known for improving information retention, no low-cost and widely accepted, effective solution exists to digitize such a pen trace. The challenge is to accurately track the pen's trajectory without an external reference system while allowing unrestricted freedom of pen movement across a surface. We propose an innovative solution that combines a digital pen, advanced artificial intelligence algorithms, and adaptive AI techniques to reconstruct the digital trace of handwriting. Our approach integrates hardware development, focusing on a sensor-equipped pen, with software innovations to optimize trajectory reconstruction and processing in real time using an embedded AI. This work aims to advance the state-of-the-art in automated trace reconstruction of handwriting, enabling a seamless connection between traditional handwriting on paper and capturing the trace digitally.**

Keywords*— *Handwriting Trace Reconstruction, IMU signals, Deep Learning, Embedded Systems, Hardware/Software Co-Design

I. Introduction

Handwriting on paper and on digital devices largely coexists without significant overlap, requiring a choice between writing on paper or using a digital pen with a tablet. Each one has its advantages. On the one hand, digital handwriting is widely used in many applications, such as note-taking or annotating documents allowing to share and store what is written. In learning handwriting, it allows a wide range of applications and exercises with appropriate feedback on the written trace produced by the student. On the other hand, writing on paper offers simplicity and availability. In addition, various research studies have demonstrated the significant benefits of hand writing in developing learning and know ledge acquisition for children, as compared to typing on a keyboard [1], [2].

With advancements in electronic pens and intelligent software, these developments in equipment and systems offer the opportunity to bridge the gap between the two worlds and to benefit from what each one offers. In this paper, we present a complete hardware-software system combining an electronic pen with compatible firmware and drivers that can work with any support such as screen, tablet and paper. In this work, we focus on dealing with handwriting on paper which is the main challenge and the major novelty. Our proposal is based on: 1) a digital pen, called Digipen, developed by STABILO International GmbH and which is equipped with inertial sensors to track pen movements; 2) various strategies to deal with handwriting signals and the use of a deep learning AI system to reconstruct handwriting traces from the inertial sensors; 3) AI algorithm optimization techniques to compress the network and embed the technology into the pen. By distributing system complexity between hardware and software, our proposal aims to enable efficient and fast AI execution.

II. Digipen: A digital pen for trajectory tracking

A growing fraction of electronic tablets is sold together with a matching electronic pen. However, a multitude of mutually incompatible technologies exists. Especially for low-cost tablets, which are attractive for schools, matching electronic pens are often not available. In addition, a clear majority of parents and teachers prefer that children learn handwriting on paper and not on a tablet. Therefore, a need exists for an electronic pen which writes on paper and can interact with all tablets.

The STABILO Digipen (Fig. 1) is a sensor-enhanced writing instrument with internal data processing capabilities and an external BLE (Bluetooth Low Energy) datalink for communication with mobile devices. It registers accelerations and rotation rates in three axes and the force on the writing tip. At the same time, it can be used as a regular pen, on regular paper. Therefore, it makes no special demands on the tablet outside of BLE connectivity and would be the ideal complement for any school tablets.

Development was started with the goal of measuring handwriting proficiency in children. Next, we add handwriting recognition using software adapted from the well-known Kaldi toolkit, originally designed for speech recognition [3]. This matches the movement patterns with prerecorded patterns, which works well for experienced writers. However, since the pattern matching relies on overlearned movements, and children who are still in the process of learning handwriting cannot yet produce

writing movement patterns with the required reliability, a different method of processing the data acquired by the pen is needed. Therefore, the development to reconstruct the pen trace only from inertial data by means of machine learning was started. Note that [4] previously proposed a similar pen based on IMU sensors but without the need to reconstruct the writing trace, which requires additional developments to acquire training samples.



Fig. 1. Experimental pen with piggybacked battery.

To acquire the needed training data, we develop a special version of the Digipen to record both pen trace and inertial data at the same time. The electronics of Wacom inking pens is combined with an experimental version of the Digipen PCB, to be used on a tablet with Wacom's recording layer and software. With a sheet of paper on the tablet, writing on paper and recording both inertial data and the pen trace as ground truth at the same time can be accomplished. Still, different sampling rates as well as delays in the transmission and processing of the BLE data required re-synchronizing the recordings of the Wacom software and the inertial data. This is achieved by re-sampling and using the onset of the tip force signal of both recording types as a clapperboard (Fig. 2).

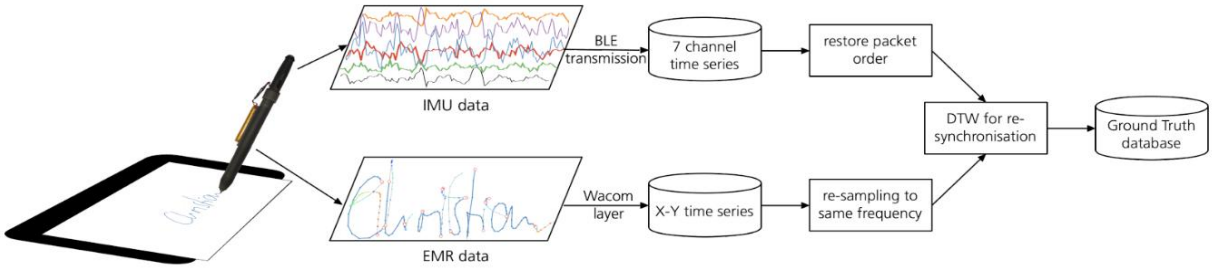


Fig. 2. Processing pipeline to construct a ground truth database. Following [11], a Dynamic Time Warping (DTW) alignment is applied to match IMU signal time frames with time frames from the tablet trajectory (EMR data).

III. Deep neural networks for handwriting trace reconstruction from IMU sensors

Handwriting trace reconstruction from online handwriting has become a key focus in many systems, including pen-and-tablet setups and camera-based configurations. While offline to online transition can be a solution [5], it often results in the loss of complex ductus, the intricate dynamics of strokes that characterize handwriting. Many tools able to capture digital handwriting traces are based on the use of a specific pen on a tablet, and it does not allow writing on paper. Camera-based tools like the Anoto pen [6] digitize handwriting using specialized microdot-patterned paper, but this approach is limited by the need for specific physical paper and the cost of the stylus. In this context, Inertial Measurement Unit (IMU) sensors are gaining traction as an affordable, low-power and versatile alternative, despite challenges associated with their inherently noisy signal. IMUs have demonstrated compatibility with digital tablets and even traditional pen-and-paper handwriting setups.

Here, we address the task of reconstructing handwriting traces from IMU signals by writing with the Digipen on paper. We explore various strategies to deal with handwriting on paper. Inspired by previous works [10], [11], we benefit from a complete pipeline including advanced preprocessing and a mixture-of-experts' approach based on deep learning models to reconstruct digital handwriting traces by writing on paper.

A. Related Works

Only a few papers have explored handwriting trace reconstruction from IMU signals. Wehbi et al. [7] extend the monowriter approach proposed by Ott et al. [8] to deal with multiple writers. This approach, based on deep learning techniques, uses handwriting on a tablet from an earlier version of the STABILO Digipen (version 5.5). In this work, the authors introduce a Convolutional Neural Network architecture designed for handwriting trace reconstruction, with linear interpolation to align sensor data with tablet-based ground truth for network training.

More recently, [11] proposed a complete pipeline for handwriting trace reconstruction from Digipen data. This pipeline relies on a new pre-processing including a Dynamic Time Warping alignment to align sensor data with tablet-based ground truth for network training and a neural network model inspired by Temporal Convolutional Networks (TCNs). Based on this work, we previously proposed a mixture of experts' approach that considers the specific characteristics of the signals [10]. In this work, two expert models are designed, one dedicated to touching strokes and one for pen-up movements. In another work, a domain adaptation technique based on a Domain-Adversarial Neural Networks has been explored to process children's handwriting [12], exhibiting varying levels of graphomotor gesture development, including differences in fluency and speed. Until now, all the experiments were done by writing on a tablet.

B. Dealing with handwriting on paper

Input signals vary significantly depending on whether the handwriting is captured on paper or on a tablet screen, due to differences in surface properties and friction. Paper typically offers higher friction, providing more resistance to the motion of a pen and causing additional noise in the IMU signals compared to handwriting on a tablet. This strongly impacts the performance of a handwriting reconstruction system trained only on data from tablets.

To reconstruct the handwriting trace, we focus on the Mixture-Of-Experts approach based on Temporal Convolutional Networks which has been successfully used with input signals on data acquired on tablets using the Digipen [10]. As in [10], we will refer to this model as MOE-CI. Here, we propose to compare different strategies to study the impact of noise and the differences between handwriting signals from tablet or paper. First, to benefit from the collection of tablet data [10], [11], we apply a Kalman filter to the signals from the pen on tablet in an attempt to reduce the noise and then evaluate the reconstruction using an MOE-CI model trained on tablet data. The second strategy consists in training an MOE-CI model from scratch on data acquired on paper using the dual acquisition process discussed above. We will compare those two strategies to the reference model from [10], trained only on handwriting acquired on a tablet and evaluate the paper data on it.

IV. Hardware-aware Neural Network Compression

The designed MOE-CI network enables handwriting trajectory reconstruction from IMU-data, however, to perform real-time inference on resource-constrained devices such as electronic pens, the neural network must be compressed for the target hardware. The current MOE-CI model is composed of two expert models. In total it requires 3.6 MB of ROM, 288 KB of RAM and 90 MFLOP for a single inference, whereas a typical system-on-chip suitable for an electronic pen has maximum 1 MB of ROM and 256 KB RAM in total.

To meet these hardware constraints, we conducted the search for the resource optimized yet still capable expert TCN models using our hardware-aware Neural Architecture Search (NAS) approach (Fig. 3). For that, we introduced expressive search space for the TCN-based models, which allows to vary the number of convolutional filter blocks of each individual layer besides the standard parameters of the TCNs such as kernel size, dilation factor, number of stacks and number of layers per stack. To efficiently sample architectures from the resulting search space, we apply the NASWOT zero-cost proxy [9] instead of full NN training, since the NASWOT score correlates with the final accuracy of the trained network and at the same time can be evaluated quickly. The search of architectures was performed using the methods of multiobjective Bayesian optimization with NASWOT score maximization and FLOP count minimization goals. We trained several architectures from the resulting Pareto front and evaluated them in hardware. The best discovered architecture requires 10,952 parameters against 467,452 of the initial TCN expert model and as little as 74 KB RAM and 66 KB ROM, pertaining comparable trajectory reconstruction accuracy and allowing to deploy these networks onto the tiny microcontroller of the Digipen.

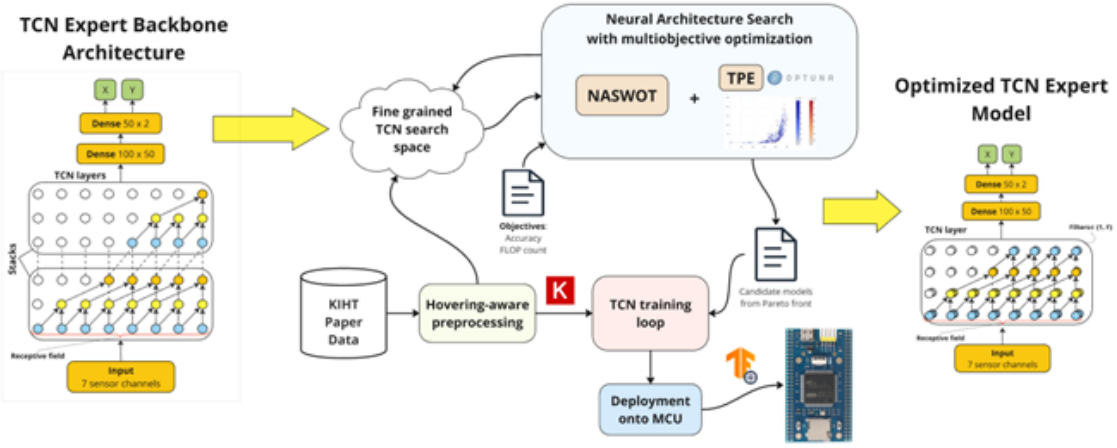


Fig. 3. Proposed hardware-aware network optimisation concept.

V. Results

Following [11], we evaluate the models and the various strategies by computing the Fréchet distance. We present quantitative and qualitative results in Table I and Fig. 4. The smaller the mean Fréchet distance, the more accurately the reconstruction matches the ground truth. Signal differences depending on the acquisition medium (tablet or paper) lead to a considerable degradation in accuracy, underscoring the model's difficulty in generalizing to differently noisy data. This is illustrated by applying an MOE-CI model trained only on tablet data compared to a specific training on data from paper.

TABLE I. MEAN FRÉCHET DISTANCE (↓) BETWEEN RECONSTRUCTED TRAJECTORY AND GROUND TRUTH.

Model	MOE-CI			Compressed MOE-CI
Acquisition medium	Tablet screen	Tablet screen	Paper	Paper
Filter	×	*Kalman filter*	×	×
Fréchet Distance	0.566	0.542	**0.429**	0.738

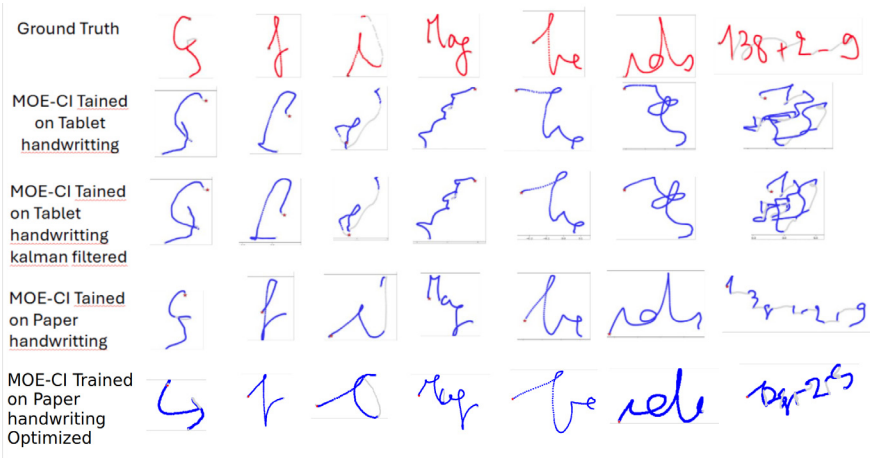


Fig. 4. Examples of ground truth traces and reconstructed traces for different models and by applying the various strategies proposed in this work.

To address this, the first proposal is to apply a Kalman filter to the signals, aiming to reduce noise. Although the Kalman filter succeeded in smoothing the data and mitigating high-frequency noise, the improvement in mean Fréchet distance is marginal and insufficient to bridge the gap between tablet and paper data. A noise reduction process may not be sufficient to account for the inherent differences in writing media.

The second strategy consisting in training a model on paper data significantly reduced the mean Fréchet distance, yielding notably better performance in reconstruction. This result underlines the importance of optimizing a model on data from the target acquisition medium, and adapting it to the specific characteristics and noise of the input data.

The hardware-aware neural network compression proposal results in a significant increase in the mean Fréchet distance, showing a degradation in trajectory reconstruction. However, the trace reconstruction when the pen touches the paper is quite close to that of the non-optimized model. It is primarily the next stroke repositioning in case of pen-up movement that suffers from the compression. This may be related to the expert model dedicated to the pen-up movement, which is trained on the entire sequence (including the touching parts) to enhance its reconstruction capabilities. In the compressed version, this could provide information that is not useful for pen-up movement reconstruction and affect performance. This highlights, however, that the compressed model remains effective in handling data written on paper.

ACKNOWLEDGMENT

This work was funded by the German Federal Ministry of Education and Research (BMBF) under grant number 01IS21097A (KIHT). This project is financed by the KIHT French-German bilateral ANR project and these four partners, IRISA (France), KIT (Germany), Learn&Go (France) and STABILO (Germany). This work was performed using HPC resources from GENCI IDRIS (Grant 2021-AD011013148)

REFERENCES

[1] P.A. Mueller and D.M. Oppenheimer, “The pen is mightier than the keyboard: Advantages of longhand over laptop note taking,” Psychological science, 25(6), 1159-1168, 2014

[2] S. Oviatt, A. Cohen, A. Miller, K. Hodge, and A. Mann, “The impact of interface affordances on human ideation, problem solving, and inferential reasoning,” ACM Transactions on Computer-Human Interaction (TOCHI), vol. 19, 2012

[3] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, and others, “The Kaldi speech recognition toolkit,” in IEEE 2011 Workshop on Automatic Speech Recognition and Understanding, 2011.

[4] F. Lunardini, D. Di Febbo, M. Malavolti, M. Cid, M. Serra, L. Piccini, A. L. G. Pedrocchi, N. A. Borghese, and S. Ferrante, “A smart ink pen for the ecological assessment of age-related changes in writing and tremor features,” IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–13, 2020.

[5] M. M. Elmokhtar, T. Lelore and H. Mouchère, “Point to segment distance DTW for online handwriting signals matching”, in 12th International Conference on Pattern Recognition Applications and Methods (ICPRAM), Portugal: SCITEPRESS - Science and Technology Publications, pp. 850–855, 2023.

[6] C. Liao, F. Guimbretière, K. Hinckley, and J. Hollan, “Papiercraft: A gesture-based command system for interactive paper”, in ACM Transactions on Computer-Human Interaction (TOCHI), 14(4), 1-27, 2008.

[7] M. Wehbi, D. Luge, T. Hamann, J. Barth, P. Kaempf, D. Zanca and B.M. Eskofier, “Surface-free multi-stroke trajectory reconstruction and word recognition using an imu-enhanced digital pen”, Sensors, 22(14), 5347, 2022.

[8] F. Ott, D. Rugamer, L. Heublein, B. Bischl and C. Mutschler, “Joint classification and trajectory regression of online handwriting using a multi-task learning approach,” In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 266-276, 2022.
[9] J. Mellor, J. Turner, A. Storkey, and E. J. Crowley, “Neural Architecture Search without Training,” In International conference on machine learning (PMLR), pp. 7588-7598, June 2021.
[10] F. Imbert, E. Anquetil, Y. Soullard, R. Tavenard, “Mixture-of-experts for handwriting trajectory reconstruction from IMU sensors”, Pattern Recognition, Volume 161, 2025.
[11] W. Swaileh, F. Imbert, Y. Soullard, R. Tavenard and E. Anquetil, “Online handwriting trajectory reconstruction from kinematic sensors using temporal convolutional network,” International Journal on Document Analysis and Recognition (IJDAR), 26(3), 289-302, 2023.
[12] F. Imbert, R. Tavenard, Y. Soullard and E. Anquetil, “Domain Adaptation for Handwriting Trajectory Reconstruction from IMU Sensors,” In International Conference on Document Analysis and Recognition, pp. 3-11, Automatically Domain-Adapted and Personalized Document Analysis (ADAPDA) workshop. Cham: Springer Nature Switzerland, 2024.

Evaluation of the Precision of Markerless Motion Capture for Pediatrics Upper-Limb Kinematics

A.Naaim[1], J.Rozaire[1,2], L.Mercier[3], S.Duprey[1], M. Begon[4]
[1]*Univ Lyon, Univ Gustave Eiffel, Univ Claude Bernard Lyon 1, LBMC UMR_T9406, Lyon, France*
[2]Centre Médico-Chirurgical de Réadaptation des Massues Croix-Rouge française, Lyon, France
[3]Laboratoire d'Innovation ouverte - École de technologie supérieure, Montréal, Canada
[4]S2M lab, Université de Montréal, Montréal (Qc), Canada
alexandre.naaim@univ-lyon1.fr

Abstract -This study assessed the precision of markerless motion capture technology for upper limb kinematics in paediatric populations. Eight typically developing children (mean age 4.7 ± 1.2 years) performed six daily tasks including drawing while being recorded simultaneously by a reference optoelectronic system (10 cameras, 120 Hz) and a markerless setup comprising eight synchronized RGB cameras. Reflective markers were placed on key anatomical landmarks according to ISB recommendations and measured with the optoelectronic systems. 2D key points were detected using the RTMPOSE neural network on the video obtained from the RGB cameras. A weighted direct linear transformation algorithm was then applied to triangulate these points into 3D coordinates. The accuracy was quantified by comparing triangulated joint centres against reference data across various camera configurations for triangulations. Results indicate that median errors ranged between 10 and 20 mm for most joints. Configurations using as few as three cameras produced performance comparable to the full eight-camera setup, with only a marginal increase in non-reconstructed points (approximately 2–4%). These findings underscore the potential for selecting optimal camera configurations tailored to specific neural networks and highlight the need for further analysis of the minimum confidence threshold on triangulation accuracy. Our comprehensive database will serve as a valuable resource for detailed evaluations of each step in the markerless tracking process and will be completed with 12 additional participants. The database will allow evaluating future markerless methods or retraining models using additional key points.

Keywords—Motion analysis, Markerless, Upper Limb, Paediatrics, Drawing

I. Introduction

Assessing to the performance of children with neurological or musculoskeletal disorders during their upper limb daily activities (e.g., drawing, writing, or Assisting Hand Assessment [1]) may benefit from more automated motion analyses. Optoelectronical systems are expensive and require highly qualified personnel and time-consuming participant setup and post-processing. Qualitative assessment by therapist is subjective, less sensitive to improvement, and time consuming. Markerless motion capture technology is an emerging and promising tool for analysing complex upper limb movements, such as writing and drawing, in various populations [2]. Systems based on synchronized RGB cameras are less expensive than optoelectronical systems and less likely to distract children performing the tasks. Moreover, the setup is faster.

While several reliability studies [2-4] have been achieved in the analysis of lower limb motion based on markerless technologies, the biomechanical accuracy of markerless systems for the upper limb remains underexplored, particularly within paediatric populations. Current approaches typically rely on multiple calibrated RGB video inputs, where 2D key points are detected — often using neural networks trained on large datasets like COCO [6] — and then triangulated into 3D coordinates using methods such as direct linear transformation or its weighted version [7]. Upper limb validation studies are still limited to single tasks and focus only slightly on paediatrics population. This study aims to bridge this gap by creating a comprehensive database of RGB videos of typically developing children and 3D reference kinematics to evaluate each step of the markerless motion analysis process. As a first step, the present study seeks to determine the optimal setup for capturing upper limb movements. Ultimately, this research will contribute to the establishment of biomechanically accurate markerless models, facilitating improved analysis of upper limb behaviours in paediatric populations.

II. Material and Method

A. Participants

Eight participants (2 boys, 6 girls) were recruited for this study, with a mean age of 4.7±1.2 years, mean weight of 18.6 ±4.5 kg, and mean height of 104±10 cm. Only typically developing children were included. The study was approved by the ethics committee of the Université de Montréal and informed consent was obtained from the participants and their parents.

Participants were asked to perform 6 daily living activities (drawing, playing with modelling clay, combing their hair, hand to back, and mimicking eating and cutting food). Participants were wearing a tank top to facilitate the placement of reflective markers.

An experimenter described the tasks to the children, who then reproduced them on their own, behind curtains to make sure only one person was seen by the cameras.

B. Material

Participants were recorded using two camera systems (Fig, 1, left). Ten optoelectronic cameras (120 Hz), considered as the reference, measured the positions of the reflective markers placed on the children's upper limbs. Eight calibrated and synchronized RGB cameras (60 Hz) were used to film the children, they were divided into five groups according to their position with respect to the participant (A: front, B: 45° front, C: lateral, D: back 45°, E: back) (Fig. 1, right). The optoelectronic system was calibrated using the Motive software procedure, while the RGB camera calibration (intrinsic and extrinsic) was performed using the Pose2Sim toolbox [8]. The 3D points were systematically measured with the optoelectronic cameras and reprojected onto all 2D video of one acquisition as a verification.

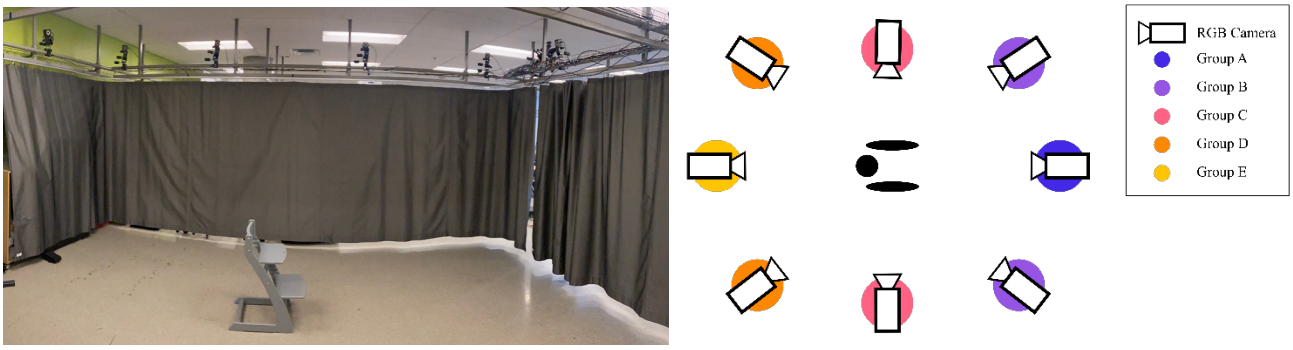


Fig. 1. Setup (left) and orientation of RGB (top view; right) cameras relative to the participants with color-coded groups of cameras (A to E).

C. Marker placement

Prior to data acquisition, markers (⌀1 mm) were placed on the following anatomical landmarks according to ISB recommendations for joint kinematics [9]: Incisura Jugularis (IJ), Processus Xiphoid (PX), spinous process of the 7th cervical vertebra (C7), spinous process of the 10th thoracic vertebra (T10), most dorsal point on the acromioclavicular joint (AC), medial (EM) and lateral epicondyles (EL), ulnar (US) and radial styloid (RS), and the heads of the 2nd (M2) and 5th metacarpals (M5). The glenohumeral joint centre was defined using the regression method by Rab et al. [10]. The centre of the elbow, wrist, and hand were defined as the midpoints of EL and EM, US and RS, and M2 and M5, respectively.

D. Key point detection on 2D images

One neural network model was used for key-point identification: RTMPOSE [11]. This model detected the positions (and their confidence values) of the shoulder joint centre (SJC), elbow joint centre (EJC), wrist joint centre (WJC), and two points on the hand, which the middle approximated to the midpoints of the 2nd and 5th metacarpal heads (HMJC). The hip joint centres (HJC) were also tracked since the chosen pose estimator has no key points on the thorax to define a trunk system of coordinates. All key-points detection was done using Mmpose which is an open-source toolbox for pose estimation based on PyTorch that includes state-of-the-art deep learning models.

E. Statistics

The effect of each camera group on the triangulation was quantified by calculating 3D accuracy of the data. Triangulation was performed for 26 combinations of camera groups (AB, AC, AD, ..., ABCDE). For each joint centre, the percentage of points' loss (i.e., when the cameras' set could not reconstruct the point) and the distance from the marker-based reference data were quantified. Triangulation was performed using a weighted direct linear transform algorithm with joint pose estimation likelihood [7] implemented in the Pose2Sim toolbox. The default 0.5 confidence threshold was used to discard or not a value for the triangulation. For both analysis, errors from the left and right sides of each joint were combined. Descriptive statistics were generated.

III. Results

Overall, median errors typically range between 10 and 20 mm for most joints (Fig. 2), although the shoulder (SJC) tends to exhibit higher deviations. As expected, configurations with less camera tend to exhibit higher maximal errors but they never exceed 2 times the errors obtained with all cameras. Configurations employing at least four cameras, and even some configurations using three cameras, generally produce errors comparable with those observed when using the full set of cameras. However, this benefit comes at the expense of a slightly higher proportion of non-reconstructed points, generally within 2–4%.

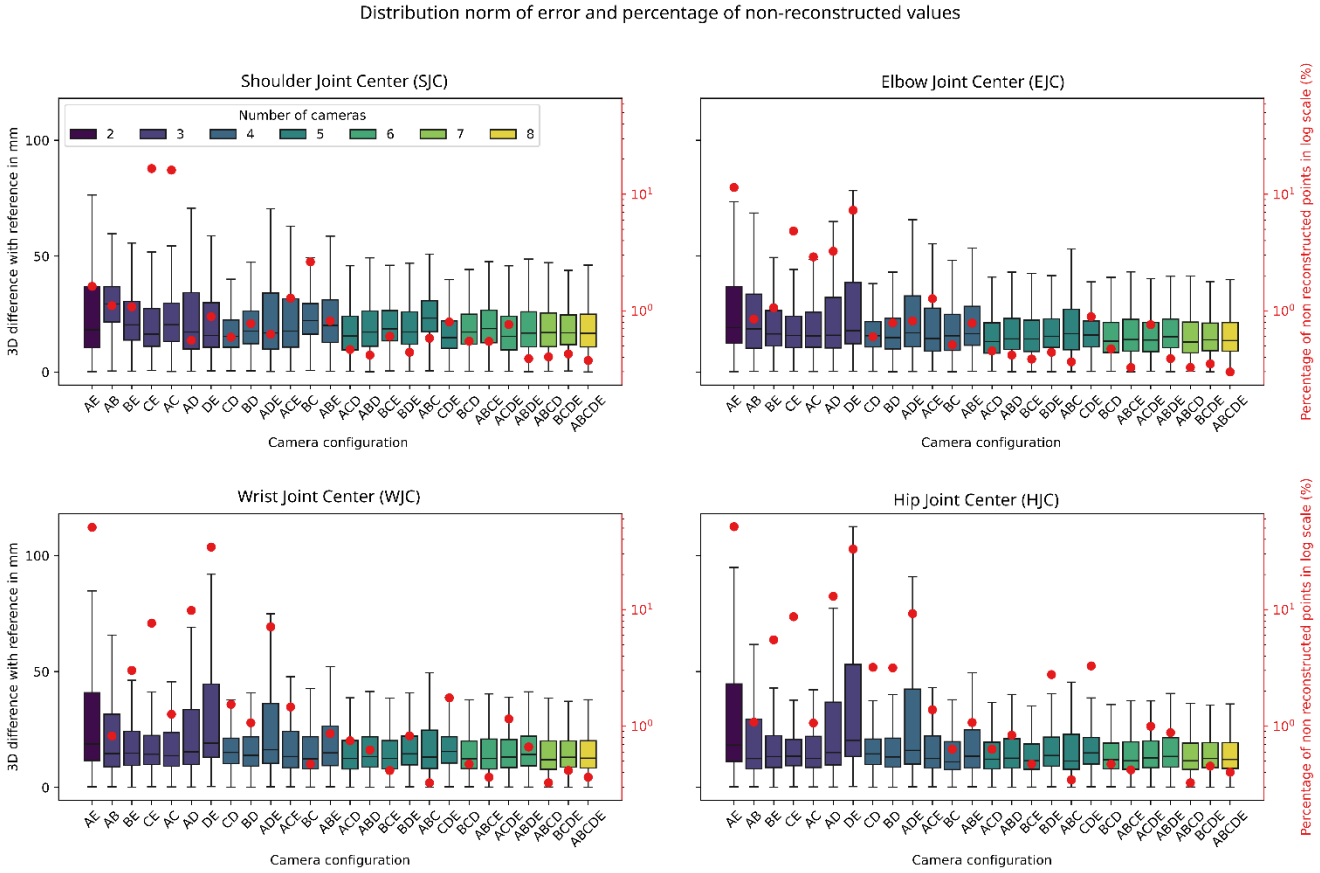


Fig. 2. 3D accuracy for each joints for the best models for all joints with all configurations of camera.

IV. Discussion

The present study's findings align with existing literature, such as the work by Lahkar et al. [12], who reported errors of 23, 31, and 18 mm for the shoulder, elbow, and wrist, respectively, using Theia MarkerLess (Kingston, Ontarion) system during adult boxing motions. Our results are slightly more accurate than in the Lahkar et al. studies, with median errors ranging from 10 to 20 mm. This seems consistent with adult vs children size.

A notable finding is that a minimal setup using only three cameras (*e.g.*, AC configuration: in front and on left and right sides of the participant) yields a slight increase in error and a data loss of approximately 2–4% relative to the full eight-camera configuration. This suggests a three-camera configuration is nearly as effective in capturing the arm, forearm, and hand positions. Future investigation should explore whether interpolation or other corrective methods can mitigate these data losses without significantly compromising the overall integrity of the kinematic measurements.

However, it must be noted that it is mandatory to also record the thorax positions to study the thoraco-humeral kinematics. In the current setup, the thorax is indirectly defined by the hip and shoulder joints, the only key points detected by the COCO whole body data set [6]. It indicates the need to incorporate hips directly within the field of view, which can be problematic when participants are seated, a very likely position when studying upper limb in paediatric pathological populations such as cerebral palsy. In such conditions, cameras in the back of the participant are essential. A configuration like CD, which offers a back view, appears to provide the optimal compromise by facilitating hip reconstruction while maintaining robust upper limb data collection. It is worth noting that children seated on a bench and pathological population would require a more stable seat, affecting the findings of the present study and the need to define the thoracic system of coordinated differently.

The current implementation uses the RTMPOSE model for 2D key-point detection, a choice that comes with its own limitations. RTMPOSE is predominantly trained on datasets such as COCO, which contain a limit number of points labelled from maily adult postures. This potential bias could contribute to inaccuracies in detecting the anatomical landmarks in children. To mitigate these issues, retraining or fine-tuning the model using paediatric-specific data could be beneficial. It would allow to eventually enhance accuracy and allow to add additional points.

A critical aspect that could affect the precision of 3D reconstruction is the minimum confidence threshold applied during triangulation. The default threshold of 0.5, used in this study, may not be optimal for all camera configurations and joint detections. Change in the confidence threshold could lead to significant differences in reconstruction errors. An approach would involve plotting the 2D error against the confidence scores for each camera orientation and joint. This analysis would allow for the determination of optimal thresholds tailored to each specific camera group, thereby minimizing error propagation during triangulation. Future work should focus on establishing a systematic method for threshold optimization, which could include iterative testing and validation against a reference system, ensuring that the balance between data retention and reconstruction accuracy is finely tuned. Such an approach would further allow to optimize camera setup for measuring upper limb kinematics using markerless technology.

The comprehensive database created in this study serves as a valuable resource for evaluating each step of the markerless tracking procedure, enabling the community to determine the optimal algorithms and camera configurations for specific neural networks. Moreover, the database opens for future developments. Its extensive dataset could be used to increase the number of points detected by training new neural network. Such work should investigate, model retraining for improved thorax estimation and additional point for the elbow and the wrist.

V. Conclusion

This study has provided an evaluation of the accuracy of markerless motion capture technology for upper limb kinematics in paediatric populations. Our findings indicate that, while an eight-camera setup yields robust data, a reduced configuration—using as few as three cameras—can produce comparable with only a marginal increase in non-reconstructed points. While studying the kinematic accuracy is important, translating these evaluations into clinically meaningful parameters remains an important next step.

Acknowledgment

All authors are members of the LIA EVASYM. This work was partly funded by a grant of the *Société de Biomécanique*.

References

[1] L. Krumlinde-Sundholm, M. Holmefur, A. Kottorp, et A.-C. Eliasson, « The Assisting Hand Assessment: current evidence of validity, reliability, and responsiveness to change », *Dev. Med. Child Neurol.*, vol. 49, nº 4, p. 259-264, avr. 2007, doi: 10.1111/j.1469-8749.2007.00259.x.

[2] R. M. Hansen, S. L. Arena, et R. M. Queen, « Validation of upper extremity kinematics using Markerless motion capture », *Biomed. Eng. Adv.*, vol. 7, p. 100128, juin 2024, doi: 10.1016/j.bea.2024.100128.

[3] A. Chaumeil, B. K. Lahkar, R. Dumas, A. Muller, et T. Robert, « Agreement between a markerless and a marker-based motion capture systems for balance related quantities », *J. Biomech.*, vol. 165, p. 112018, mars 2024, doi: 10.1016/j.jbiomech.2024.112018.

[4] R. M. Kanko, E. K. Laende, E. M. Davis, W. S. Selbie, et K. J. Deluzio, « Concurrent assessment of gait kinematics using marker-based and markerless motion capture », *J. Biomech.*, vol. 127, p. 110665, oct. 2021, doi:10.1016/j.jbiomech.2021.110665.

[5] T. E. McGuirk, E. S. Perry, W. B. Sihanath, S. Riazati, et C. Patten, « Feasibility of Markerless Motion Capture for Three-Dimensional Gait Assessment in Community Settings », *Front. Hum. Neurosci.*, vol. 16, p. 867485, juin 2022, doi: 10.3389/fnhum.2022.867485.

[6] S. Jin *et al.*, « Whole-Body Human Pose Estimation in the Wild », 23 juillet 2020, *arXiv*: arXiv:2007.11858. doi: 10.48550/arXiv.2007.11858.

[7] Y. I. Abdel-Aziz et H. M. Karara, « Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates in Close-Range Photogrammetry », *Photogramm. Eng. Remote Sens.*, vol. 81, nº 2, p. 103-107, févr. 2015, doi: 10.14358/PERS.81.2.103.

[8] D. Pagnon, M. Domalain, et L. Reveret, « Pose2Sim: An open-source Python package for multiviewmarkerless kinematics », *J. Open Source Softw.*, vol. 7, nº 77, p. 4362, sept. 2022, doi: 10.21105/joss.04362.

[9] G. Wu *et al.*, « ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion—Part II: shoulder, elbow, wrist and hand », *J. Biomech.*, vol. 38, nº 5, p. 981-992, mai 2005, doi:10.1016/j.jbiomech.2004.05.042.

[10] G. Rab, K. Petuskey, et A. Bagley, « A method for determination of upper extremity kinematics », *Gait Posture*, vol. 15, nº 2, p. 113-119, avr. 2002, doi: 10.1016/S0966-6362(01)00155-2.

[11] T. Jiang *et al.*, « RTMPose: Real-Time Multi-Person Pose Estimation based on MMPose », 3 juillet 2023, arXiv: arXiv:2303.07399. doi: 10.48550/arXiv.2303.07399.

[12] B. K. Lahkar, A. Muller, R. Dumas, L. Reveret, et T. Robert, « Accuracy of a markerless motion capture system in estimating upper extremity kinematics during boxing », *Front. Sports Act. Living*, vol. 4, p. 939980, juill. 2022, doi: 10.3389/fspor.2022.939980.



Session

Robotic & Arts

Drawing, Mind, and Movement: Possible Directions for Technological Exploration

Seymour Simmons III, Ed.D.
Professor of Fine Arts Emeritus, Winthrop University, Rock Hill, SC 29730
seymoursimmons@gmail.com

Abstract—Responding to recent advances in "fMRI-compatible tablet technologies" and their capacity to "approximate naturalistic … drawing behaviors inside a scanner" (Lin, *et al,* 2021), this paper discusses the significance of research on drawing and movement as reflections of "mind in motion." After introducing the research, I will outline three drawing experiments, each involving distinct forms and degrees of movement, proposing them as topics for further study. The aim of this discussion is to highlight drawing's potential to engage a diversity of cognitive, perceptive, affective, and sensory-motor processes which are of particular importance to life and learning in the 21st century.

Keywords—*Drawing, fMRI, Gesture, Kinesthetics*

I: Introduction

Drawing, in its fluidity, focus, and search for structure, vividly exemplifies the integration of mind and body with implications in and beyond the arts, to include all four STEM subjects: science, technology, engineering, and mathematics. Thus, cognitive psychologist, Barbara Tversky, in *Mind in Motion* (2019), cites a variety of drawing activities as evidence that thinking in all such spheres evolved from, and remains grounded in, movement. Elsewhere, Tversky (2011, p. 15) puts the kinds of thinking that are done in and through drawing in a larger context, saying: "Thinking is hard. When thoughts overwhelm the mind, the mind puts them into the world, and has since antiquity. We use fingers, tallies, abacuses, computers to count and calculate. We gesture maps and routes in the air or draw them in sand, in stone, on paper, on screens … These cognitive artifacts, externalizations of thought, expand the mind."

This article seeks to support such claims based on recent research using fMRI tracking technologies to link movements of the hand with movements in the brain. To begin, I will describe the technological advances in the experiments that prompted this paper and indicate the relevance of such research in the teaching of drawing. Then, anticipating further advances in such technology, I will propose three additional experiments in drawing to document more complex and distinct patterns in brain and body. To conclude, I will discuss some implications for such research to support a "renaissance" in drawing instruction to help meet human needs in the Digital Age.

II: Capturing Mind and Movement in Drawing

As an artist, art educator, and researcher specializing in drawing (Simmons & Winer, 1977; Simmons, 2021), I have followed with interest research and speculation on the relationship between drawing and the brain. Initially, this interest was prompted by findings that associated drawing with functions located in the right cerebral hemisphere, as addressed in psychologist Howard Gardner's *The Shattered Mind* (1977) and in Betty Edwards' best-selling instructional text, *Drawing on the Right Side of the Brain* (1979). In both cases, the main takeaway for me at least was that drawing was indeed a cognitive capacity. However, since the kinds of drawing addressed in these books were located on the right side of the brain, the thinking involved was assumed to be intuitive rather than logical, creative more than critical.

While I valued this body of research, I also realized that it did not encompass drawing as a whole and so was restrictive pedagogically. On one side, it ignored rigorous forms of "academic" drawing as well as the logic and mathematics required for technical drawing in architecture and design. On the other, it shortchanged more physical, active drawing exemplified by gestural sketches of the figure, animals, etc. Ultimately, drawing reduced to a right-brain phenomenon could not accommodate the notebooks of Leonardo da Vinci: his anatomical accuracy; his use of rigorous perspective applied to compositions, buildings, and machinery; his intensely felt renderings from nature: trees and storm clouds, birds in flight, horses and warriors in battle; even the methodical way he made his lines and marks. (Isaacson, 2017)

Given such concerns, I sought studies that documented cerebral activity in real time doing different kinds of drawings. Therefore, after being informed about the 2025 IGS conference, I was pleased to find an article co-authored by Simon J. Graham, one of the Keynote Speakers, entitled "Tablet Technology for Writing and Drawing during Functional

Magnetic Resonance Imaging: A Review." (2021) Among other things, the article highlighted a "computerized tablet system technology, which, through use of a touch-sensitive surface and accompanying stylus, permits effective fMRI studies of writing and drawing behaviour." The article continues with "a review of various fMRI applications where tablet systems have been utilized, and a discussion of how this technology and research are likely to progress in the future." Although the use of the tablet allows more freedom of drawing movement and observation of the drawing in process than would be possible if it took place entirely inside an fMRI tube, the drawing process still appears to be relatively restrictive, but it is easy to imagine continued advances in the technology to eventually allow more free and naturalistic ways of drawing. It is with such hopes that I propose the following experiments based on actual drawing lessons and related activities.

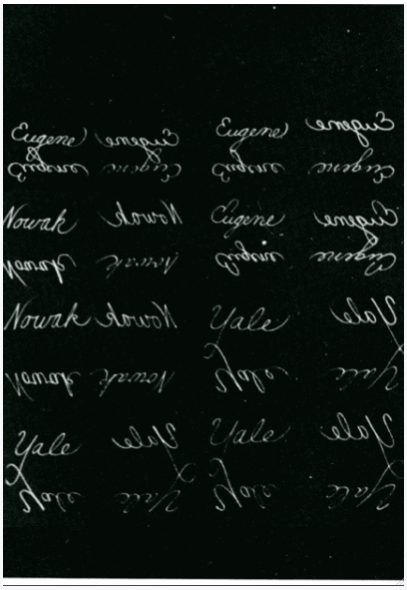
III: Three Experiments Proposed

In the hands of a competent drawer, the following experiments might well be possible using the touch sensitive screen and stylus employed in the article, "Tablet Technology for Writing and Drawing." More ideal circumstances would allow the participant a fuller range of motion and upright posture from sitting at a table or standing at an easel, not to mention on the floor or a wall. These not only allow for more freedom of movement but invite the artist to step back to monitor the work in progress by comparing the drawing to the model or idea at each stage of the process. Likewise, as lessons, these activities would benefit from teacher demonstrations, feedback from peers, and the freedom of being able to stop and start all over again whenever one wants.

IV: Drawing and Writing

In the spirit of experiments noted above on both writing and drawing, this first experiment treats writing as a form of drawing. It was the first exercise of the semester in William Reimann's "Freehand Drawing" course at Harvard University. The course assumed no prior experience and was open to all students, undergraduate and graduate. The course was grounded in classes taught by Reimann's own teacher, Josef Albers, at Yale University. Albers had previously been a student and teacher at the Bauhaus and also taught at Black Mountain College in North Carolina.

This first lesson was designed, in Albers' words (1969, p. 12), to "get [students] in touch with their instrument -- that is, to do as musicians do when they practice." In drawing, that involved becoming conscious of how hands, arms, and shoulders function together. It began by having students "draw" their own signatures. At first, they drew in the air over and over to make conscious what they usually did habitually. But once students could write their names while being more consciously aware, the next tasks had them push those skills in other directions, *literally:* drawing their signatures in reverse, then upside down, and finally reversed and upside down. (Figure 1)



Source: The Josef and Anni Albers Foundation; Also, from *Search versus Re-Search:* Courtesy of the Watkinson Library, Trinity College, Hartford, Connecticut.

Reimann wrote about this assignment on an instruction card for his students, entitled: "On the sense of motor-kinesthesis." "One's physically internal awareness of being in motion plays a vital role in support of the fine discriminations made by eye. The first awareness exercises, writing in the air, then in the air with the eyes closed, seek to place such awareness momentarily at the center of attention. Through the movement of the arms each person certifies the process of experiencing a sensation first, and later connecting it to an idea, a purpose, and a specific goal… [B]ut most importantly, the learning process is shown to extend beyond passive use of the 'head' alone… It is the coordination of eye, mind, and hand that is of necessity one of the most highly keyed events in drawing."

Having experienced this lesson as a student, and later, as a graduate teaching assistant, having helped others to solve the problem of drawing a signature accurately in different orientations, there is some serious "headwork" involved along with the motor skills that I also hope would be evident in the fMRI scans. It involves treating the signature, no longer primarily as a name, but as a design element that can be reversed, drawn upside-down, etc. In the final form, the four signatures reflect one another in a design format called "four-fold mirror symmetry." What happens to most students, including myself, is initially a sense of frustration at the difficulty, which evolves as they find a way to do it. This transformation, with its cognitive, motor, and emotional components, could, I believe, be recorded in the scan.

V: Seeing, Thinking, Moving, and Drawing

The following experiment is based on a presentation for a conference at the *Institut d'études avancées de Paris*: "Space-Time Geometries and Movement in the Brain and in the Arts." The description of this international conference highlights "the importance of understanding how the brain perceives and represents space, time and movement, and how it plans and controls our bodily movement and actions." Speaking on "The Role of Neural Circuitry in Skilled Drawing," Emilio Bizzi, a neuroscientist from the Massachusetts Institute of Technology (MIT), gave a close description of how the brain functions when transferring an image in the mind or eye to the actions needed to produce a skilled drawing, such as a Renaissance portrait study. As Bizzi explained, the production of such masterly drawings rests on a number of neural circuits organized hierarchically with feedback loops at every level of the central nervous system.

He then went on to describe the neurological mechanisms underlying the formation of spatial-temporal patterns of motor activity involved in drawing and briefly touched upon the formation of motor memories that result from the intense practice. Specifically, he pointed out that "neural signals representing skilled learning originate in the frontal areas of the brain: dorsal lateral prefrontal cortex, Motor Cortex, SMA (Supplementary Motor Area), and Parietal cortex." From there, signals go into the subcortical area, a specific region in the basal ganglia where they are segmented into 'chunks' or 'syllables' of shorter action commands that descend into the spinal cord shaped in such a way that they can produce movements. Regarding learning, he said that "practice leads to improvement in the quality of drawing by establishing memory traces through a hierarchically organized neural circuitry." Bizzi concluded by saying that fMRI technology now can make visible the more subtle and mysterious factors that lead to art, but that these studies are still waiting to be done. In reference to the present paper, this is the potential I hope to witness one day, given the new technologies as noted above. In the meantime, as an experiment, this process could take on several forms.

The ideal, depending on a relatively high skill level, would be drawing a living person from direct observation. However, similar patterns might be produced by copying a master drawing or drawing from a photograph. Extreme alternatives would be to draw a caricature or comic book character based on one of the above. While the former requires strict observation of the subject, the latter engages imagination and exaggeration to bring out unique features. In each case, because the subject is another person, every interpretation should involve a mix of thought and feeling

VI: Movements of the Body

After the presentation above, architect Paul Andreu offered an alternative to Bizzi's sequence of events. Rather than starting with an "image in the mind," Andreu said that his ideas often emerged through the act of drawing itself. Tversky cites Andreu in *Mind in Motion,* referring to him as "the visionary architect of Charles de Gaulle Airport in Paris…" Responding to Tversky's question, "'Where do you begin?' 'Inside,' he said. Meaning that he begins with the user/s

experience: you need to be uplifted by the space you are in and you need to be guided effortlessly to where you need to go" (p. 268). Discussing her own work, Tversky spoke about the reasons she got into this kind of research:

"When I was a graduate student, the reason I got into this whole thing, language had the hegemony in thinking, [but] it seemed to me that space came first evolutionarily and in the life of every individual. Understanding space has its own logic. We're first acting in space and language is based on that, not vice versa. [In drawing,] it becomes a conversation between the eye and the hand doing the action and what appears on the page... We think thinking happens between the ears, but the conversation with gesture is spatial motor." (quoted in Simmons, 2019, p. 5)

VII: Movements of the Body

The second example recalls another exercise, called "gesture drawing" that Tversky might call "bottom up." Gesture drawing is best explained in the classic drawing text, *The Natural Way to Draw,* by Kimon Nicolaides first published in 1941. Gesture drawing exemplifies a 'natural,' experiential, way to draw. In this case, seeing involves visual scanning guided by kinesthetic/proprioceptive awareness and empathy. As such, gesture is concerned less with the shape or even the volume, but more with the "energy" of the subject, including even inanimate objects. Likewise, while contour or outline drawing is done slowly and "painstakingly" in order to accurately capture an edge, Nicolaides says gesture drawing must be done fast and "furiously" moving up the middle, then out and back in again, as illustrated in Figure 2.



Figure 2. Source: Seymour Simmons III, "Orgue de barbarie", 2017

The instructions, following Nicolaides, are as follows: "The model is asked to take a very active pose for a minute or less . . . As the model takes the pose . . . you are to draw, letting your pencil swing around the paper almost at will, being impelled by the sense of action you feel. Draw rapidly and continuously in a ceaseless line, from top to bottom, around

and around, without taking your pencil off the paper. Let the pencil roam, reporting the gesture. You should draw, not what the [subject] looks like, not even what it is, but what it is doing." (ibid., p. 14–15)

As an experiment, this process could certainly be carried out and perhaps carried out more easily with a pressure sensitive graphic tablet in a scientific study by giving the person a choice of drawing without looking at the tablet at all. I do this most of the time when I am starting to draw again after a long absence. I know the results will not be "good drawings," but sometimes they are surprisingly evocative of the subject matter, more so than a closely observed study.

Leonardo, in fact, recommended a similar process, in which artists constantly carry a small sketchpad around with them to capture people in motion, people expressing emotions, people arguing with each other, and other everyday experiences. (Kelen, 1990, p. 89) Eventually, perhaps there will be monitoring systems that can document the brain involved in such tasks, but for the time being, participants can simply be aware of their own internal responses to external events, and how they feel about the experience: doubt, discomfort, sympathetic and antipathetic emotions, etc.

Conclusion

Each of these examples, drawing as a visual language in handwriting, drawing in portraiture, and figure drawing, offers opportunities for investigation with the advanced imaging technologies noted at the outset. Each in their own ways exemplify the wedding of mind and body in movement, with drawing as a cognitive domain not localizable in one side of the brain or the other but potentially engaging the whole brain and the person as a whole. Such a skill as exemplified by Leonardo and promoted from the Renaissance through the Industrial Revolution (Simmons, 2021), has practical, experiential, and educational importance, not only for those aspiring for careers in the visual arts, but for everyone especially in an age dominated by visual media, immersed in visual information, misinformation, and *dis*information.

In the conclusion of their article, Lin *et al,* included the following as a possible future benefit: "Utilizing the digitizing feature of the tablets, enhanced behavioural recording and performance metrics become possible to examine behavioural kinematics efficiently with greater detail than is possible with traditional paper-based tasks." Drawing, as indicated above, can enable us to see more accurately and sensitively, as well as to judge more objectively how well or how poorly we observe, thereby challenging us to be more visually attuned to others while becoming more sensitive to our own thoughts and feelings. Therefore, drawing processes monitored in the ways outlined above may have the potential not just to help us improve our drawing performance, but to make us more aware both outwardly and inwardly. This indeed would be a gift of technology: quite the opposite of our everyday technological distractions and deceptions.

Acknowledgment

With thanks to Frederic Fol Leymarie for his encouragement and friendship over the years.

References

[1] J. Albers. Search Versus Research, Hartford: Trinity College Press, 1969.
[2] B. Edwards. Drawing on the Right Side of the Brain. Los Angeles, Tarcher, 1979.
[3] Z. Lin, F. Tam, N. W. Churchill, T.A. Schweizer, & S. J. Graham, Tablet technology for writing and drawing during functional magnetic resonance imaging: a review. Sensors, 21(2), 401. https://doi.org/10.3390/s21020401, 2021.
[4] W. Isaacson, Leonardo da Vinci. New York: Simon and Schuster, 2017.
[5] E. Kelen, Leonardo da Vinci's Advice To Artists. Philadelphia: Running Press, 1990.
[6] K. Nicolaides, The Natural Way to Draw. Boston: Houghton Mifflin, 1941.
[7] S.Simmons & M. S. Winer, Drawing: The Creative Process. Englewood Cliffs: Prentice Hall, 1977.
[8] S.Simmons, Drawing in the Digital Age: Observations and Implications for Education" *Arts, 2019 8,33.* 2019.
[9] S. Simmons, The Value of Drawing Instruction in the Visual Arts and Across Curricula. New York: Routledge, 2021.
[10] B. Tversky, "Obsessed by Lines," in Thinking Through Drawing: Practice into Knowledge, A. Kantrowitz, A. Brew, M. Fava, Eds. New York: Teachers College, 2011.
[11] B. Tversky, Mind in Motion, New York: Basic Books, 2019.

Predicting Differential Aesthetics in Offline Handwritten Signatures

Cristian Rodriguez, Moises Diaz, Miguel A. Ferrer
iDeTIC, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain
cristian.rodriguez.110@alu.ulpgc.es, {moises.diaz, miguelangel.ferrer}@ulpgc.es

***Abstract*— This paper proposes a predictor that quantifies the differences between the aesthetics of two signatures by developing a differential aesthetic score. Although aesthetics is inherently subjective, we aim to make this comparison objective by using a weighted sum of geometric features such as rectangularity, slope, and quadrant distribution, among others. To determine the optimal feature weights, we collected 2,332 responses from over 100 volunteers who judged the relative aesthetic of signature pairs executed by the same signer. These responses served as the ground truth for training machine learning regression models to derive the feature weights. We then evaluated the effectiveness of the learned differential aesthetic score, which achieved an accuracy, recall, precision and F1-score of 79.72%, 90.72%, 81.48% and 85.85% respectively in predicting subjective aesthetic preferences between two signatures. To validate the proposed differential aesthetic score, we conducted two evaluations: (1) a comparison of signatures from a public dataset and a celebrity signature dataset, with the latter proving to be more aesthetically pleasing, and (2) an analysis of multiple signatures from the same individual to identify the most aesthetically pleasing specimen. These results suggest that, despite individual differences in perception, certain quantifiable factors contribute to the perceived beauty of handwritten signatures.**

Keywords*— *Offline signatures aesthetics, geometric features, survey, applications, machine learning

I. Introduction

Graphonometry has approached the study of handwriting and signatures from different perspectives. Some research correlates handwriting and signatures with personality traits or features influenced by cultural, social and psychological factors [1]. The aesthetics of the handwritten signatures have been recognized as a key element in fields such as politics, business, culture and the arts [2]. Public figures often seek distinctive and visually appealing signatures to represent their personal image.

This paper explores the evaluation of handwritten signature aesthetic through its geometric features, aiming to develop a descriptive equation for selecting the most visually appealing option. While research specifically addressing signature aesthetics is limited, related studies on handwriting provide valuable context. Prior work has explored how handwriting traits relate to personality and aesthetic perception [1], examining both macro aspects (e.g., layout, spacing) and micro elements (e.g., word construction) [3], as well as the visual quality of Chinese calligraphy using neural network-based features [4].

Based on these studies, this research hypothesizes that signature aesthetics can be objectively evaluated by comparing the geometric features of signatures. Since aesthetics can be seen as a comparative quality, the proposed system is based on peer comparison. To evaluate this hypothesis, a survey[1] was conducted in which participants compared pairs of signatures based on aesthetic preference. Using the majority votes as ground truth, 23 geometric features were extracted from offline signatures and machine learning models were trained to develop a differential aesthetic score equation.

To validate the differential aesthetic score equation, two evaluations were performed: (1) the aesthetic appeal of a personal database of celebrity signatures[2] compared to the BiosecurID database [5], which includes signatures of individuals who did not prioritize aesthetics, and (2) the selection of the most aesthetically pleasing signature among multiple iterations of signatures of the same individual. This will demonstrate that the system has the potential to provide personalized advice to individuals seeking to enhance the visual appeal of their signature. This could be particularly valuable for VIPs, CEOs, politicians, and other public figures who use their signatures in official or corporate contexts where they are often seen by a wide audience.

II. Methodology

In this section, we describe the process used to develop the signature aesthetic scoring system. To do so, we start by defining the differential aesthetic score equation (ΔAS). Being $f_{i,1}$ the characteristics of the offline signature 1 and $f_{i,2}$ the characteristics of the offline signature 2, we define Δf_i as $f_{i,1} - f_{i,2}$ and ΔAS as the sum of all the features, as shown in (1):

$$\Delta AS = \omega_1 \Delta f_1 + \omega_2 \Delta f_2 + \cdots + \omega_{23} \Delta f_{23} \tag{1}$$

When $\Delta AS > 0$, signature 1 is considered more beautiful, and when $\Delta AS < 0$, signature 2 is considered more beautiful. The higher the value of $|\Delta AS|$, the greater the aesthetic difference between the two signatures. The constants ω_i represents the weight of each feature Δf_i, which determines their relative importance. To define this equation, two key components are needed: the values of the features (f_i) and the corresponding weights (ω_i).

[1] https://rb.gy/yq01sl [Accessed: Apr. 29, 2025]
[2] https://rb.gy/8umqa4 [Accessed: Apr. 29, 2025]

A. Geometric features

The geometric features have been designed to try to describe key aesthetic elements that contribute to the overall visual appeal of a signature, taking into account the principles of balance, symmetry, and harmony that are central to Gestalt design principles [6]. These features, inspired by both human perception and previous research on calligraphic aesthetics [4], form the foundations for developing the differential aesthetic score equation. As an illustrative example, Fig.1 and Fig.2 show some of these features applied to a randomly selected signature from the celebrity database. Below is a detailed description of the 23 extracted features:

- Rectangularity (f_1). Ratio between the perimeter of the convex envelope and the perimeter of the bounding box, defined as $f_1 = \frac{P_{envelope}}{P_{bounding_box}}$. Being the convex envelope the smallest convex shape that encloses all the pixels of the signature, which can be visualized as the red line in Fig.1. The perimeter of the bounding box is the outline of a minimum rectangle that can enclose the signature. This feature is illustrated in Fig.1 as the blue box.
- Inclination of the signature (f_2, f_3). Obtained by fitting a straight line to the pixels of the signature by a first degree polynomial fit, which determines the coefficients ($p = (a, b)$) of the polynomial ($Y \approx aX + b$). These coefficients are calculated by solving the least squares problem with a being the slope of the signature (f_2) and b the pixels of intersection of the line with the vertical axis (f_3). Specifically, f_2 is calculated as the absolute value of a, while f_3 is the absolute value of b normalized by the width of the bounding box. This is represented as the green line in Fig.1.
- Center of gravity proportions (f_4, f_5). Normalized coordinates of the center of gravity within the bounding box. They are computed as $f_4 = \frac{1}{W}\sum x_i$, and $f_5 = \frac{1}{H}\sum y_i$, where W and H are the width and height of the bounding box, and x_i, y_i correspond to the coordinates of individual pixels of the signature. It is shown as the green dot in Fig. 1.
- Pixel density (f_6). This feature measures the density of the signature pixels in relation to the area of the convex envelope: $f_6 = \frac{N}{A_{\text{envelope}}}$ where N is the number of pixels in the signature.
- Quadrant distribution ($f_8 - f_{11}$) and elastic mesh ($f_{12} - f_{17}$). The first four features, $f_8 - f_{11}$, describe the distribution of signature pixels relative to the center of gravity, divided into four quadrants. The remaining features describe how the signature is distributed within its bounding box by dividing both the horizontal and vertical ranges of the bounding box into 4 equal segments, referred to as the elastic mesh. These segments are then used to calculate the relative distribution of pixels both horizontally, $f_{12} - f_{14}$, and vertically, $f_{15} - f_{17}$, as shown in Fig.2. These features are calculated as the proportion of pixels in each quadrant or segment, relative to the total number of signature pixels.

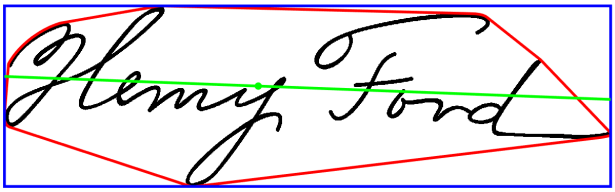


Fig. 1: Rectangularity, gravity center and slope.

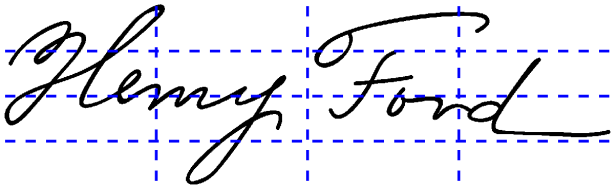


Fig. 2: Elastic mesh.

- Space intra-characters (f_{18}). Amount of empty space within the characters of the signature. It quantifies how open or compact the writing is by evaluating the filling of these inner spaces. $f_{18} = \frac{|C_{gap}|}{C_{gap}+Ctotal}$, where C_{gap} is the number of pixels added to fill inner spaces in the signature, and C_{total} is the total number of pixels in the original signature image.
- Projections on the axes ($f_{19} - f_{22}$). Variance of the sum of the pixel values along the different axes. The projections include the horizontal axis (f_{19}), vertical axis (f_{20}), 45° axis (f_{21}, obtained by rotating the image 45 degrees) and 135° axis (f_{22}, obtained by rotating the image 135 degrees).
- Signature symmetry (f_{23}). Comparison of the number of pixels on the left and right sides of its center of gravity.

B. Survey design

To assess the aesthetic appeal of handwritten signatures, we used a subset of the BiosecurID [5] dataset, selecting signature numbers 4 and 16 from each user. Signature 4 was chosen to represent a point at which the user was already familiar and comfortable with the digitizer and stylus, while signature 16 was chosen as the last enrolled signature in the sequence. These pairs were presented to participants, who selected the more aesthetically pleasing option. A total of 107 anonymous participants contributed to 2,354 pairwise comparisons. Surveys were intentionally brief and user-friendly to promote broad engagement, and no personal data were collected to avoid incurring in any legal aspect associated with the privacy of the participants. Each pair was independently evaluated, and the average agreement among participants was 69.43%; notably, only one signature pair received unanimous agreement, while eight showed complete disagreement. A majority voting method was used to determine the preferred signature for each pair. These responses, along with the corresponding subset of signatures from the BiosecurID dataset, were used to fit the features of the differential aesthetic score equation.

C. Model training and weights adjustment

To determine the optimal weights for the differential aesthetic score (ΔAS), we evaluated several machine learning models: linear regression, support vector machine (SVM), k-nearest neighbors (kNN), neural networks (NN), and random forest (RF) using MATLAB's Classification Learner app with default hyperparameters. The task was a binary classification problem based on differences in geometric features, applying 5-fold cross-validation and SMOTE to correct class imbalance.

The generalized linear model (fitglm), corresponding to a linear regression classifier, achieved the best result with an accuracy, recall, precision, and F1-score of 79.72%, 90.72%, 81.48%, and 85.85% respectively. In comparison, the kNN model reached 74.13%, 61.86%, 99.98%, and 76.43%; the SVM model 78.32%, 89.69%, 80.56%, and 84.88%; the RF model 82.52%, 79.38%, 93.90%, and 86.03%; and the NN model 79.72%, 86.03%, 84.54%, and 85.28%. Although random forest achieved a slightly higher accuracy, fitglm was selected because it offered the best balance between recall and precision, which were prioritized over accuracy alone to ensure robust and reliable performance. The resulting feature weights (Table I) highlighted pixel density and symmetry as positive contributors to perceived aesthetics, while imbalance in quadrant distributions and diagonal projections had a strong negative effect, aligning with Gestalt principles favoring balance and harmony.

TABLE I: Feature weights divided between positive and negative.

Positive feature	**Weight**	**Negative feature**	**Weigth**
Pixel rate	$w_6 = 4.2616$	1st column elastic mesh	$w_{15} = -8.9409$
Intersect Point	$w_3 = 3.2455$	Y projection	$w_{21} = -6.3421$
Pixels quadrant 3	$w_{10} = 2.3213$	45º projection	$w_{20} = -6.2923$
Intra-character	$w_{18} = 1.5777$	135º projection	$w_{22} = -5.3341$
Symmetry	$w_{23} = 1.3472$	Pixel per half	$w_7 = -4.1547$
Rectangularity	$w_1 = 1.3254$	2nd row elastic mesh	$w_{13} = -3.4147$
Gravity Center Y	$w_5 = 0.5724$	Gravity Center X	$w_4 = -3.3042$
X projection	$w_{19} = 0.4498$	Pixels quadrant 1	$w_8 = -2.6537$
-		Slope	$w_2 = -1.9102$

* Missing features (w_9, w_{11}, w_{12}, w_{14}, w_{16} and w_{17}) have values smaller than 0.0001.

III. Validation

To validate the aesthetic scoring equation, two evaluations were conducted. The first involved a comparative analysis between 100 celebrity signatures and 132 signatures from the BiosecurID database. Each celebrity signature was tested against all BiosecurID samples, resulting in a preference percentage. 64% of the celebrity signatures had a preference rate above 50%, suggesting that the predictor tends to favor celebrity signatures. This suggests that public figures may place greater importance on the visual quality of their signatures because of their public exposure. However, this interpretation should be approached with caution, as not all celebrities prioritize aesthetics in their signatures. Nevertheless, these signatures have become publicly recognized as they represent each individual's personal image.

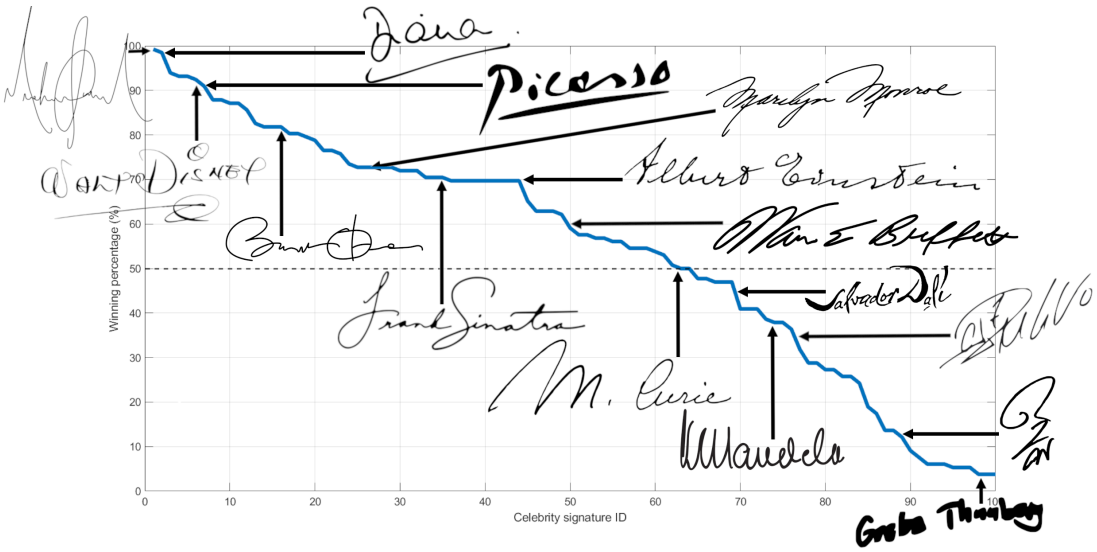


Fig. 3: Comparison of celebrities vs. BiosecurID signatures. This figure also shows some examples of celebrity signatures at different levels of the line. For privacy reasons, only openly published celebrity signatures were displayed.

The second evaluation is intended to simulate an automatic query system that chooses the most aesthetic signature from a set produced by a person. This situation may be useful for CEOs, VIPs, politicians who use the signature corporately in their associations or companies. To simulate this query functionality, the three authors of this study each provided 16 sample signatures. These signatures were compared following a winner plays next tournament format, comparing in pairs and the winner advances to compare with the next in chronological order (from 1 to 16). The results are shown in Fig. 4, where some sample signatures are shown and the one selected by the system as the most aesthetic is highlighted. This second evaluation illustrates the system's ability to help users refine their personal signatures by presenting them with the most objectively aesthetic version and allowing them to improve it.

The second evaluation is intended to simulate an automatic query system that chooses the most aesthetic signature from a set produced by a person. This situation may be useful for CEOs, VIPs, or politicians who use their signature corporately in professional contexts. To simulate this query functionality, the three authors of this study each provided 16 sample signatures. These were compared following a winner plays next tournament format, where the selected signature in each comparison advanced to the next round in chronological order (from 1 to 16). The results, shown in Fig. 4, highlight the signature chosen by the system as the most aesthetically pleasing. Our evaluation objectively illustrates the model's ability to consistently select one preferred signature from a set. A weakness of this method is that another survey would be required to confirm whether our system's objectivity aligns with human aesthetic perception.

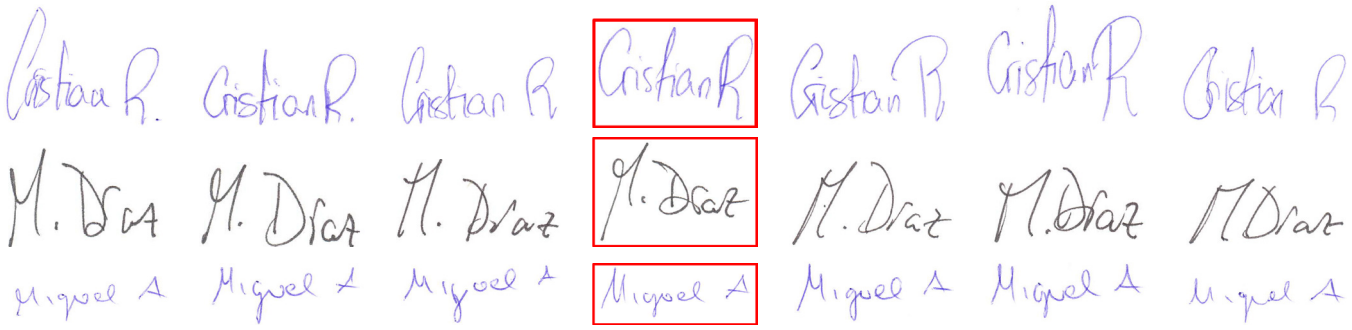


Fig. 4: A set of signatures from three individuals, with the signature that is most preferred by the predictor being highlighted in a red box.

IV. Conclusion

The study evaluates the aesthetics of offline signatures by developing a differential system based on the comparison of different geometric features. Using a survey, machine learning models were trained to determine the optimal weighting of the features, with linear regression achieving the best results (79.72% accuracy). The differential aesthetic score equation was validated through two evaluations: comparing celebrity and non-celebrity signatures, and selecting the most aesthetic signature from individual sets. Findings suggest that quantifiable geometric traits, such as pixel density, symmetry, and spatial distribution, consistently influence aesthetic perception. While the system shows potential as a personalized tool for signature refinement, especially for professionals and public figures, further human-based validation would be beneficial to confirm the model's recommendations in subjective scenarios. Overall, the approach offers a novel, objective framework for analyzing and improving signature aesthetics.

Acknowledgment

This research was supported by the PID2023-146620OB-I00, funded by MICIU/AEI 10.13039/501100011033, and the European Union's FEDER program and the CajaCanaria and la Caixa (2023DIG05).

References

[1] K. Chaudhari and A. Thakkar, "Survey on handwriting-based personality trait identification," *Expert Systems with Applications*, vol. 124, pp. 282–308, 2019.

[2] A. Hakami, S. Alminhali, and A. Jamal, "Personality identification based on handwritten signature," *2nd International Conference on Emerging Trends in Information Technology and Engineering, ic-ETITE 2024*, 2024.

[3] A. Majumdar, P. Krishnan, and C. V. Jawahar, "Visual aesthetic analysis for handwritten document images," *Proceedings of International Conference on Frontiers in Handwriting Recognition, ICFHR*, vol. 0, pp. 423–428, 7 2016.

[4] R. Sun, Z. Lian, Y. Tang, and J. Xiao, "Aesthetic visual quality evaluation of chinese handwritings," in *International Joint Conference on Artificial Intelligence*, 2015.

[5] J. Galbally, M. Diaz-Cabrera, M. A. Ferrer, M. Gomez-Barrero, A. Morales, and J. Fierrez, "On-line signature recognition through the combination of real dynamic data and synthetically generated static data," *Pattern Recognition*, vol. 48, no. 9, pp. 2921–2934, 2015.

[6] B. Smith, "Gestalt theory: An essay in philosophy," in *Foundations of Gestalt Theory* (B. Smith, ed.), pp. 11–81, Philosophia, 1988.

Differentiable Rasterization of Minimum-Time Sigma-Lognormal Trajectories

Daniel Berio
Goldsmiths, Univ. of London
London, United Kingdom
daniel.berio@gold.ac.uk

Sylvain Calinon
Idiap Research Institute
Martigny, Switzerland
sylvain.calinon@idiap.ch

Réjean Plamondon
Ecole Polytechnique de Montréal
Montréal, Canada
rejean.plamondon@polymtl.ca

Frederic Fol Leymarie
Goldsmiths, Univ. of London
London, United Kingdom
ffl@gold.ac.uk

***Abstract*— We present an adaptation of the sigma-lognormal model to generate and fit smooth trajectories in conjunction with a differentiable vector graphics (DiffVG) rendering pipeline and with parameter selection driven by a minimum-time smoothing criterion. This approach enables the incorporation of the "Kinematic Theory of Rapid Human Movements" into modern image-based deep learning systems. We demonstrate its utility through various applications, including fitting handwriting trajectories to an image and generating trajectories using guidance from a large multimodal model.

Keywords*— *Sigma-lognormal model, Differentiable Vector Graphics, DiffVG, curve generation.

I. Introduction and background

Recent advances in automatic differentiation [1] and differentiable rendering [2] have enabled the development of numerous methods that use complex image-based costs to optimize parameters for 2D and 3D parametric shape primitives. In the 2D computer graphics domain, differentiable vector graphics (DiffVG) rasterization [3] supports most primitives in the Scalable Vector Graphics (SVG) format, allowing gradients to propagate through the transformation of these primitives into pixels of an image. This capability has facilitated various methods for image vectorization [4], stylization and abstraction [5, 6], as well as sketch generation [7, 8, 9].

One limitation of the SVG format for representing drawings, sketches and handwriting is its reliance on cubic Bézier curves. With this representation, longer curves are constructed by joining multiple polynomial segments at their endpoints, a representation that struggles to capture both the degree of continuity and intrinsic variability that characterize human movement. In previous work [10, 11], we demonstrated that the sigma-lognormal ($\Sigma\Lambda$) model provides a valid alternative to Bézier curves to model handwriting and drawing traces. Additionally, we showed that the model's formulation can be adapted to enable an intuitive user interface similar to those typically used in computer-aided curve design.

In this paper, we show that with minimal adaptations, a similar approach can be integrated with a DiffVG engine. This integration is particularly valuable in combination with various deep-learning techniques, enabling a wide range of potential applications using modern image generation and processing methods to drive the creation of $\Sigma\Lambda$ trajectories. We demonstrate two practical examples combining DiffVG with a minimum-time smoothing criterion [12]: (i) an approach that infers $\Sigma\Lambda$ parameters based on image similarity, similarly to existing methods [10, 13, 14], and (ii) a method for generating handwriting-like trajectories using a pre-trained multimodal (text and image) model [9].

II. Method

On the basis of the "Kinematic Theory of Rapid Human Movements" [15] we describe a planar trajectory using a weighted-displacement formulation of the $\Sigma\Lambda$ model [16]. The model describes the planar evolution of a handwriting trajectory through the space-time superposition of a discrete number of stroke primitives, each having a characteristic "bell-shaped" speed profile modeled with a two-parameter lognormal. The spatial layout of a trajectory is described with a high-level motor plan consisting of an initial position $\boldsymbol{p}_0$ followed by a sequence of "virtual targets" $(\boldsymbol{p}_1, \dots, \boldsymbol{p}_m)$ (Fig. 1a and 1b). Each virtual target describes a ballistic stroke aimed along the orientation θ_i of the vector $\boldsymbol{p}_i - \boldsymbol{p}_{i-1}$. Assuming movements are performed with rotations of the wrist or elbow, each stroke follows a circular arc geometry determined by a turning angle parameter δ_i (Fig. 1c). The time parametrized kinematics $\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t)$ of a trajectory can be computed as linear combinations

$$\boldsymbol{x}(t) = \boldsymbol{p}_0 + \sum_{i=1}^{m} \boldsymbol{x}_i(t) \quad \text{and} \quad \dot{\boldsymbol{x}}(t) = \sum_{i=1}^{m} \dot{\boldsymbol{x}}_i(t) \tag{1}$$

of m submovements with position $\boldsymbol{p}_0 + \boldsymbol{x}_i(t)$ and velocity $\dot{\boldsymbol{x}}_i(t)$.

The speed profile of each stroke follows a lognormal function (Fig. 1d), and the overall trajectory results from the superposition of multiple strokes over time. The smoothness of the trajectory depends on the activation time and overlap of consecutive lognormals (Fig. 1a and 1b). If a new lognormal starts before the previous one is zero, the superposition creates a smooth transition between the two strokes. A greater overlap results in a smoother transition.

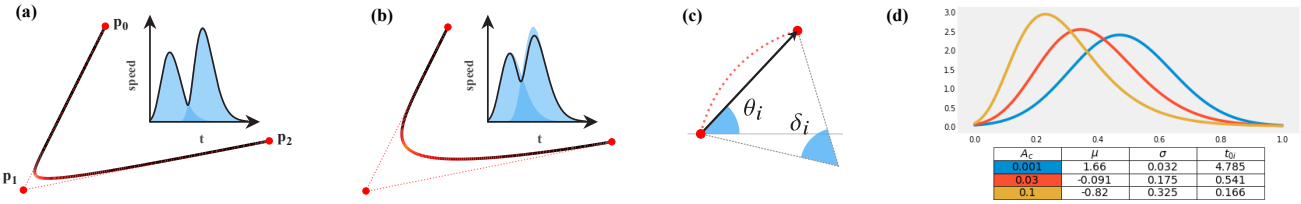


Fig. 1. **(a)** and **(b)**: The effect of different time overlaps for two lognormals, where a greater time overlap results in a smoother trajectory. The degree of overlap is indicated along the black trajectory with a red shade and the trajectory speed is filled in blue. The motor plan is made up of virtual targets (red dots) linked by their distance separation (dotted red). **(c)**: Stroke orientation θ_i and turning angle δ_i. The latter determines the internal angle of the assumed circular arc. **(d)**: Controlling lognormal shape and asymmetry using the intermediate parameter A_c and $T = 1$. The lognormals are shifted by $-t_{0i}$ for visualization, so that their onset matches at $t = 0$.

To maintain compatibility with an automatic differentiation pipeline we compute the lognormal with:

$$\Lambda_i(t) = \frac{1}{\sigma_i\sqrt{2\pi}\,\varphi(t - t_{0i})} \exp\left(-\frac{1}{2{\sigma_i}^2}(\ln(\varphi(t - t_{0i})) - \mu_i)^2\right), \tag{2}$$

where we use a rectifier linear unit $\varphi(x) = \text{ReLU}(x - \epsilon) + \epsilon$, using a small ϵ to avoid values ≤ 0 in the logarithm. The parameters t_{0i}, μ_i and σ_i, respectively determine the activation time, delay and response time of the lognormal.

To compute the curvilinear evolution of each stroke we use the following integral:

$$w(t) = \int_0^t \Lambda_i(u)du = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\ln(\varphi(t - t_{0i})) - \mu_i}{\sigma_i\sqrt{2}}\right)\right] \quad \in [0, 1], \tag{3}$$

which can be computed explicitly using the error function (erf). The velocity of each stroke is then given by:

$$\dot{\boldsymbol{x}}(t)_i = D_i\Lambda_i(t)\begin{bmatrix}\cos(\delta_i w(t) - \frac{\delta_i}{2} + \theta_i)\\ \sin(\delta_i w(t) - \frac{\delta_i}{2} + \theta_i)\end{bmatrix} \quad \text{with} \quad D_i = \left\|\boldsymbol{p}_i - \boldsymbol{p}_{i-1}\right\| \text{sinc}\left(\frac{\delta_i}{2\pi}\right)^{-1}, \tag{4}$$

where D_i determines the distance traveled by a submovement along a circular arc while the use of the normalized sine cardinal (aka sinc) function avoids numerical errors as the turning angle approaches zero. The position along each stroke can be computed explicitly as a displacement with respect to the initial position $\boldsymbol{p}_0$ and without requiring numerical integration of Eq. 4 with:

$$\boldsymbol{x}_i(t) = D_i\begin{bmatrix}\frac{\cos(\theta_{0i}+\delta_i w_i(t))-\cos(\theta_{0i})}{2\sin(\delta_i/2)}\\ \frac{\sin(\theta_{0i}+\delta_i w_i(t))-\sin(\theta_{0i})}{2\sin(\delta_i/2)}\end{bmatrix} \text{ if } |\delta_i| > \epsilon, \quad \boldsymbol{x}_i(t) = D_i\begin{bmatrix}\cos(\theta_i)w_i(t)\\ \sin(\theta_i)w_i(t)\end{bmatrix} \text{ otherwise, and } \quad \theta_0 = \theta - \frac{\pi + \delta_i}{2}.$$

III. Reparameterization

While the lognormal function works remarkably well in describing the form of human movement speed profiles [17], its parameters μ_i and σ_i influence its mode and shape while also influencing the time occurrence of the activation parameter t_{0i}. To facilitate the optimization procedure that follows, we reparameterize each submovement with: (i) a stroke duration T_i, (ii) a relative time offset Δt_i with respect to the previous stroke time occurrence and duration, and (iii) a shape parameter $A_{c_i} \in (0, 1)$, which defines the skewness of the lognormal [18]. The $\Sigma\Lambda$ parameters $\{\mu_i, \sigma_i, t_{0i}\}$ are then computed with:

$$\sigma_i = \sqrt{-\ln(1 - A_{c_i})}, \qquad \mu_i = 3\sigma_i - \ln(\frac{-1 + e^{6\sigma_i}}{T_i}), \qquad t_{0i} = t_{0i-1} + \Delta t_i \sinh(3\sigma_i). \tag{5}$$

With this formulation, the parameter Δt_i explicitly determines the smoothness of a trajectory near a virtual target, with lower values producing a greater overlap with the previous lognormal and a consequently smoother trajectory near the corresponding virtual target. Furthermore, a strictly positive Δt_i guarantees an ordering of the lognormals, which allows us to compute the end time of the trajectory with:

$$T_{\text{end}} = t_{0m} + e^{\mu_m + 3\sigma_m} \quad . \tag{6}$$

This, in turn, allows us to easily compute a minimum-time cost for the optimization procedure that follows.

IV. Differentiable rasterization

On the basis of DiffVG's support for cubic Bézier curves, we can render an approximation of the ΣΛ trajectory by leveraging the equivalent Hermite formulation of curves [19], where a cubic polynomial curve is described by two point and tangent pairs. Given the planar trajectory positions $\boldsymbol{x}(t)$ and velocities (tangents) $\dot{\boldsymbol{x}}(t)$ along with two consecutive samples t_i and $t_i + dt$, the four control points for the corresponding Bézier curve segments are simply obtained as:

$$[\boldsymbol{x}(t_i),\ \boldsymbol{x}(t_i) + 3dt\dot{\boldsymbol{x}}(t_i),\ \boldsymbol{x}(t_i + dt),\ \boldsymbol{x}(t_i + dt) - 3dt\dot{\boldsymbol{x}}(t_i + dt)] \quad ,$$

where the factors $3dt$ ensure that the control points are positioned to match the desired endpoint velocities.

We compute a fixed number of samples, n, for each rendered trajectory using $dt = T_{\text{end}}/n$. We find that, in practice, using five times as many samples as targets ($n = 5m$) gives a sufficiently good approximation for our use cases. Note that even though the number of samples is fixed, the trajectory duration changes depending on the values of Δt_i.

All the trajectory construction procedures described to this stage, including conversion to Bézier, result in a graph of differentiable operations that are seamlessly handled by standard automatic differentiation engines such as the one provided by PyTorch [1]. This, combined with the ability to rasterize an approximation of the trajectories in a differentiable manner, allows us to optimize the motor plan positions $\boldsymbol{p}_0, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_m$, the time offsets Δt_i, the turning angle parameters δ_i and, as an option, the shape parameters A_c. The latter are set with respect to a compound loss that balances trajectory smoothness, in a geometric sense, with an image-based objective. For all the examples that follow, we consider an optimization objective determined by the following compound loss:

$$\mathcal{L} = \lambda T_{\text{end}} + \mathcal{L}_{\mathbb{I}},$$

where T_{end} acts as a smoothing term by rewarding overlap between consecutive lognormals, $\mathcal{L}_{\mathbb{I}}$ is an image-based cost depending on the application, and λ is a user-defined and application-dependent scalar weight that controls the tradeoff between the two terms. In the following sections, we demonstrate two examples of variants of $\mathcal{L}_{\mathbb{I}}$ for different applications. We implement the optimization procedure in PyTorch and use the Adam optimizer for 300 steps. To aid convergence, we find it beneficial to apply a cosine annealing schedule to the learning rate and to linearly increase the value of λ from zero to its user-specified value over the first 30 steps, ensuring that the initial optimization steps are principally focused on the image-space term.

A. *Fitting trajectories to an image*

As a baseline test for our method, we test fitting ΣΛ trajectories to an image, resulting in a procedure similar to the geometry-based fitting methods described by Berio et al. [10] and Ferrer et al. [14]. We start by using tag traces from the Graffiti Analysis database [20], a dataset of graffiti tags recorded with low-cost DIY tracking devices. We render the input trajectory into a black-and-white image and then create an initial motor plan for generating a ΣΛ trajectory by applying the discrete curve evolution (DCE) polyline simplification method [21] to the input trace (Fig. 2). DCE simplifies polylines by

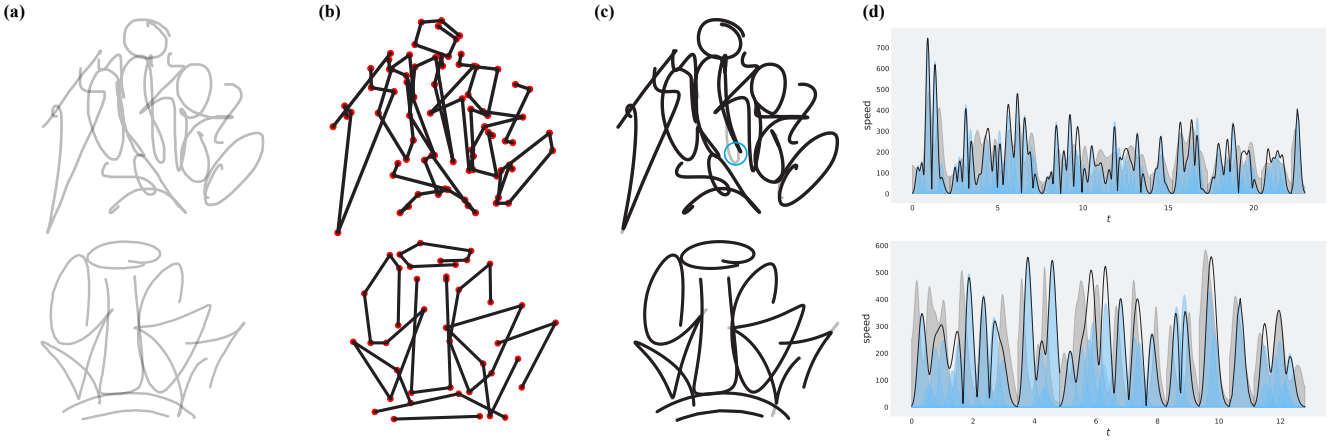


Fig. 2. Image-based trajectory reconstruction using Eq. 7. **(a)** Reconstructing an input trajectory (grey) for two graffiti tags ("JANKE" and "SUG"). **(b)** We start from the motor plan (red) and trajectory (black) computed with initial values of $\Delta t_i = 1$ and $\delta_i = 0$. The motor plan and trajectory initially match. **(c)** Trajectory optimization results . For the top row we optimize also the shape parameters A_c while for the bottom row we keep A_c fixed. A significant reconstruction error (top row, 3rd image, blue circle) occurs on the lower part of the "K", likely due to the presence of an inflection. **(d)** The generated speed profiles with the corresponding lognomals in blue and the original (Gaussian smoothed) speed profiles in grey. The latter are not taken into consideration, but the minimum time cost produces a similar number and location of peaks.

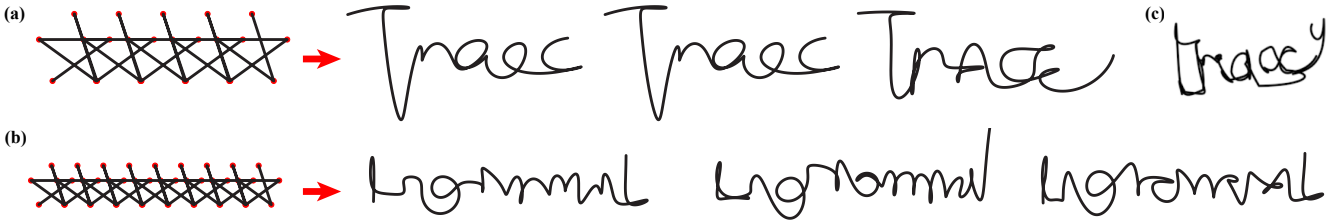


Fig. 3. CLIP-guided trajectory generation. **(a)** and **(b)**, left: We start from a motor plan (red) and a trajectory (black) with $\Delta t_i = 1$, which initially overlaps almost exactly with the motor plan. Here, the motor plan is created by concatenating multiple copies of a self intersecting polygon; however, different user-defined polygons can be used. **(a)** and **(b)**, right: Trajectories resulting from different optimizations using Eq. 8 and the target words "TRACE" and "lognormal". **(c)**, top right: Optimizing the control points of cubic Bézier curves instead of the $\Sigma\Lambda$ model parameters to generate the word "TRACE" also does not produce a fully legible output and further results in a trace with multiple discontinuous overlapping segments.

identifying significant curvature extrema, with a result that is similar to that of Berio et al. [10] and De Stefano et al. [22]. We then set $\mathcal{L}_{\mathbb{I}}$ to a multi-scale mean squared error (MSE) between the rendered image $\mathbb{I}$ and the target $\hat{\mathbb{I}}$:

$$\mathcal{L}_{\mathbb{I}} = \sum_{s \in S} \text{MSE}(\mathbb{I}_s, \hat{\mathbb{I}}_s) \ . \tag{7}$$

The MSE term consists of multiple downscaled and blurred versions $\mathbb{I}_s$ and $\hat{\mathbb{I}}_s$ of the input, where $s \in S$ represents different scaling factors applied to the images. We find that this multiscale approach improves convergence while decreasing the risk of ending in local minima. Note that the optimization infers a plausible motion from the input geometry and that is done differently from most state-of-the-art $\Sigma\Lambda$ fitting methods [13, 23], as our method does *not attempt* to reproduce the kinematics of the input.

B. CLIP-driven handwriting generation

In a second application of our method, we test an approach similar to ClipDraw [8] to guide the generation of a trace that is consistent with a given text caption or "prompt" (Fig. 3). We set

$$\mathcal{L}_{\mathbb{I}} = -\langle f(\mathbb{I}), f(c) \rangle \ , \tag{8}$$

where $\langle \cdot \rangle$ denotes the cosine similarity measured between the embeddings $f(\mathbb{I})$ of the rendered image and the embeddings $f(c)$ of the text caption. To construct $f(\cdot)$, we use the finetuned CLIPAG model [9], which has been shown to give significantly better results than the original CLIP model for image generation tasks. We use a caption "{WORD}, neatly handwritten" where {WORD} is a word we wish the trace to be similar to. Interestingly, while the generated text is not fully legible, it possesses a structure that resembles the desired word shape, suggesting that the method partially captures the high-level features of handwriting style and word formation.

V. Conclusion

We have demonstrated how $\Sigma\Lambda$ trajectories can be constructed and approximated as piecewise cubic Bézier curves through a sequence of differentiable operations. When combined with a DiffVG rendering pipeline, this approach enables the optimization of trajectory parameters using cost functions defined in image space. In this paper, we have presented two example applications: image-based trajectory fitting and CLIP-driven trajectory generation. Beyond these, numerous image-based objectives and metrics [e.g. as found in 5, 6, 24] can be used in a similar way, offering good opportunities for future research. For the fitting procedure, we rely on a correctly ordered starting trajectory. Combining our method with others that infer ordered traces from bitmap images [25] or shape outlines [26] is also an interesting avenue of future research.

Additionally, we have shown that the $\Sigma\Lambda$ parametrization allows for a straightforward computation of trajectory duration, which can be effectively used as a smoothing cost under the assumption of minimum-time as a performance criterion. The results suggest potential applications of the $\Sigma\Lambda$ model in path planning and robotics, where minimum-time is a widely used optimization objective [27].

While this paper has focused on differentiable rendering, a similar minimum-time cost function could be integrated with a geometric similarity measure for trajectory optimization. Although our initial goal was to enable creative applications of the $\Sigma\Lambda$ model, our results suggest that the proposed method is generally a useful way to reconstruct handwriting samples in terms of $\Sigma\Lambda$ parameters. In future developments, we plan to test our method on samples from handwriting signature databases with a higher sampling rate. This will allow to validate the reconstruction accuracy and efficiency of our method, while comparing it with state-of-the-art $\Sigma\Lambda$ reconstruction methods such as robust XZERO [23] and iDeLog [14].

Acknowledgements

This work has been partly funded by the EACVA (Embodied Agents in Contemporary Visual Art: eacva.org) project jointly led by Goldsmiths — under grant no. AH/X002241/1 from the UKRI/AHRC — and the University of Konstanz — under grant no. 508324734 from the Deutsche Forschungsgemeinschaft (DFG).

References

[1] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in Proc. of the 33rd Intl. Conf. on Neural Information Processing Systems, Red Hook, NY, USA: Curran Associates Inc., 2019.
[2] H. Kato, D. Beker, M. Morariu, T. Ando, T. Matsuoka, W. Kehl, and A. Gaidon, "Differentiable Rendering: A Survey," July 2020.
[3] T. Li, M. Lukáč, M. Gharbi, and J. Ragan-Kelley, "Differentiable vector graphics rasterization for editing and learning," ACM Trans. on Graphics, vol. 39, pp. 1–15, Dec. 2020.
[4] X. Ma, Y. Zhou, X. Xu, B. Sun, V. Filev, N. Orlov, Y. Fu, and H. Shi, "Towards Layer-wise Image Vectorization," in 2022 IEEE/CVF Conf. on Comp. Vision and Pattern Recognition (CVPR), (New Orleans, LA, USA), pp. 16293–16302, IEEE, June 2022.
[5] S. Iluz, Y. Vinker, A. Hertz, D. Berio, D. Cohen-Or, and A. Shamir, "Word-As-Image for Semantic Typography," Mar. 2023.
[6] Y. Vinker, E. Pajouheshgar, J. Y. Bo, R. C. Bachmann, A. H. Bermano, D. Cohen-Or, A. Zamir, and A. Shamir, "CLIPasso: Semantically-aware object sketching," ACM Trans. Graph., vol. 41, July 2022.
[7] X. Xing, C. Wang, H. Zhou, J. Zhang, Q. Yu, and D. Xu, "DiffSketcher: Text Guided Vector Sketch Synthesis through Latent Diffusion Models," Jan. 2024.
[8] K. Frans, L. Soros, and O. Witkowski, "CLIPDraw: Exploring text-to-drawing synthesis through language-image encoders," in Advances in Neural Information Processing Systems (S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds.), vol. 35, pp. 5207–5218, Curran Associates, Inc., 2022.
[9] R. Ganz and M. Elad, "CLIPAG: Towards Generator-Free Text-to-Image Generation," in 2024 IEEE/CVF Winter Conf. on Applications of Comp. Vision (WACV), (Waikoloa, HI, USA), pp. 3831–3841, IEEE, Jan. 2024.
[10] D. Berio, F. F. Leymarie, and R. Plamondon, "Kinematics reconstruction of static calligraphic traces from curvilinear shape features," in The Lognormality Principle and Its Applications in E-Security, e-Learning and e-Health (R. Plamondon, A. Marcelli, and M. Ã. Ferrer, eds.), vol. 88 of Series in Machine Perception and Artificial Intelligence, ch. 11, pp. 237–268, Dec. 2020.
[11] D. Berio, F. F. Leymarie, and R. Plamondon, "Expressive curve editing with the sigma lognormal model," in Proc. of the 39th Annual European Assoc. for Comp. Graphics Conf.: Short Papers, EG, (Goslar, DEU), pp. 33–36, Eurographics Assoc., 2018.
[12] H. Tanaka, J. W. Krakauer, and N. Qian, "An optimization principle for determining movement duration," J. of Neurophysiology, vol. 95, no. 6, pp. 3875–3886, 2006.
[13] R. Plamondon, C. O'reilly, J. Galbally, A. Almaksour, and A. Anquetil, "Recent developments in the study of rapid human movements with the kinematic theory: Applications to handwriting and signature synthesis," Pattern Recognition Letters, vol. 35, pp. 225–235, Jan. 2014.
[14] M. A. Ferrer, M. Diaz, C. Carmona-Duarte, and R. Plamondon, "iDeLog: Iterative dual spatial and kinematic extraction of sigma-lognormal parameters," IEEE trans. on pattern analysis and machine intelligence, vol. 42, no. 1, pp. 114–125, 2018.
[15] R. Plamondon, C. O'Reilly, C. Remi, and T. Duval, "The lognormal handwriter: Learning, performing and declining.," Frontiers in Psychol., vol. 4, no. 945, 2013.
[16] D. Berio, F. Fol Leymarie, and P. R., "Computer aided design of handwriting trajectories with the kinematic theory of rapid human movements," in 18th Biennial Conf. of the Intl. Graphonomics Society, 2017.
[17] B. Rohrer and N. Hogan, "Avoiding spurious submovement decompositions II," Biological cybernetics, vol. 94, no. 5, pp. 409–14, 2006.
[18] R. Plamondon et al., "A kinematic theory of rapid human movement. Part IV," Biological Cybernetics, vol. 89, no. 2, pp. 126–38, 2003.
[19] G. E. Farin, Curves and Surfaces for CAGD: A Practical Guide. Morgan Kaufmann Ser. Comput. Graph. Geom. Model., 2001.
[20] E. Roth, T. Watson, C. Sugrue, T. Vanderlin, and J. Wilkinson, "An open databse for graffiti markup language (GML) files," 2009.
[21] L. J. Latecki and R. Lakämper, "Discrete approach to curve evolution," in Mustererkennung 1998, pp. 85–92, Springer, 1998.
[22] C. De Stefano, M. Garruto, and A. Marcelli, "A saliency-based multiscale method for on-line cursive handwriting shape description," in Frontiers in Handwriting Recognition, 2004. IWFHR-9 2004. Ninth Intl. Workshop On, pp. 124–129, IEEE, 2004.
[23] C. O'Reilly and R. Plamondon, "Automatic extraction of sigma-lognormal parameters on signatures," in Proc. of 11th Intl. Conf. on Frontier in Handwriting Recognition (ICFHR), 2008.
[24] X. Xing, H. Zhou, C. Wang, J. Zhang, D. Xu, and Q. Yu, "SVGDreamer: Text Guided SVG Generation with Diffusion Model," Apr. 2024.
[25] R. Plamondon and C. M. Privitera, "The segmentation of cursive handwriting: An approach based on off-line recovery of the motor-temporal information," IEEE Trans. on Image Processing, vol. 8, no. 1, pp. 80–91, 1999.
[26] D. Berio, F. F. Leymarie, P. Asente, and J. Echevarria, "StrokeStyles: Stroke-based segmentation and stylization of fonts," ACM Trans. Graph., vol. 41, Apr. 2022.
[27] J. Bobrow, S. Dubowsky, and J. Gibson, "Time-optimal control of robotic manipulators along specified paths," The Intl. J. of Robotics Research, vol. 4, no. 3, pp. 3–17, 1985.

Articulation-Based Retrieval of Stroke Gestures

Inês Cardoso Oliveira
University of Luxembourg
Luxembourg
i.oliveira@uni.lu

Nuwan T. Attygalle
Université catholique de Louvain
Belgium
nuwan.attygalle@uclouvain.be

Réjean Plamondon
Polytechnique Montréal
Canada
rejean.plamondon@polymtl.ca

Luis A. Leiva
University of Luxembourg
Luxembourg
luis.leiva@uni.lu

Abstract

A detailed understanding of gesture articulation, i.e. how gestures are produced, can provide designers with valuable insights for developing more intuitive graphical user interfaces. Over the years, several models have been proposed to model and understand human movements, and in particular handwriting movements. Recently, Deep Learning (DL) models have become a popular approach to handwriting synthesis, being able to produce high-quality realistic samples. However, it is unclear whether these models encode articulation features. To bridge this gap, we conduct a retrieval task. As the ground truth, we employ the Sigma-lognormal ($\Sigma\Lambda$) model, a well-established descriptor of human movements. Experimental results reveal that DL embeddings and $\Sigma\Lambda$ features are mostly unrelated, suggesting a gap between these two.

Index Terms

Stroke gestures, Information Retrieval, Kinematic Theory, Deep Learning

I. Introduction

The ubiquitous adoption of touchscreen-based devices, such as smartphones and tablets, has established "stroke gestures" as a critical input modality for interacting with graphical user interfaces (GUIs). Essentially, stroke gestures are representations of a two-dimensional trajectory of one or more contact points on a sensitive surface. This type of interaction has also been extended into other domains, such as Augmented Reality glasses [4] and head-mounted displays [16], which reinforces its versatility and continued importance in the future.

Human-Computer Interaction (HCI) research has shown that stroke gestures can be employed as efficient shortcuts to access different applications or functions in a system [1]. For GUI designers, understanding gesture articulation,[1] i.e., how stroke gestures are produced, is paramount when making design decisions, for example, knowing which gestures are easier to perform, and thus easier to recall, or identifying gestures with greater variability within a population, which might not make good candidates for shortcuts. This information can be obtained through controlled studies with real users or, alternatively, by using computational models of human handwriting, saving considerable time and effort. Thus, having a effective articulation-based gesture retrieval system could provide valuable insights for the design of intuitive and user-friendly commands for GUIs.

Modeling human movements has long been a subject of interest, and many methods have been proposed; e.g., based on behavioral models [13], kinematic models [9], or minimization principles [5]. The Kinematic Theory [11] and its latest instantiation, the Sigma-Lognormal ($\Sigma\Lambda$) model [12], describes movements by modeling velocity profiles using lognormal functions, and is among the most accurate approaches so far [3]. Specifically for stroke gesture data, the Kinematic Theory has been successfully used for synthesis and analysis [6], [7]. Recently, Deep Learning (DL) approaches to handwriting synthesis have become increasingly popular, due to their capacity to generate high-quality and realistic samples [2], [8], [14]. Still, it is currently unclear whether these models encode any relevant information about articulation execution that can be leveraged to build a workable retrieval system.

In this paper, we explore whether DL models for handwriting synthesis do effectively capture articulation-based features. We analyse three different DL models: Style-disentangled Transformer (SDT) [2], Diffusion model for Handwriting Generation (DHG) [8], and DeepWriteSYN (DWS) [14]. To collect accurate articulation information, we employ the $\Sigma\Lambda$ model [12]. Our hypothesis is that similar velocity patterns mean similar gesture articulations.

II. Methodology

To evaluate the extent to which DL models for handwriting synthesis encode features related to the production process of gestures, we design and conduct a retrieval task. First, we use the $\Sigma\Lambda$ features to perform retrieval and treat these results as ground truth. We then repeat the same retrieval task using the embeddings extracted from the DL models (SDT, DHG, and DWS) and assess their ability to retrieve samples similar to the ground truth. Embeddings are lower-dimensional representations of the input data that encode complex meaningful information in a compact form. We focus on evaluating the presence of articulation information in the latent space, as it reflects the abstract features the model has learned during training.

[1]The term 'articulation' refers to the characterization of the kinematics, that is, the motion, of the hand or utensil used to produce the strokes, encompassing features such as trajectory, velocity and length.

A. Dataset

We perform experiments on $1-GDS, the most popular unistroke gesture dataset in HCI. The dataset contains 16 gesture classes executed by 10 users. Each user provided 10 samples per gesture at 3 articulation speeds (slow, medium, fast) using an iPAQ Pocket PC. Considering that a user likely executes the same symbol in a consistent way within their preferred articulation speed, for our study we use only one sample per gesture per user for slow and fast speeds. As such, the total number of samples in our experiments is 320.

B. Ground Truth Construction

The $\Sigma\Lambda$ model assumes that a complex handwritten trace can be decomposed into a series of primitives[2] that connect a sequence of virtual targets. The velocity profile of the ith primitive is described by a lognormal-shaped function, which is scaled and time-shifted by parameters D_i and t_{0_i}, and characterized by the logtime delay μ_i and the logresponse time σ_i of the neuromuscular system that produces the trace:

$$||\vec{v}_i(t)|| = D_i\Lambda(t; t_{0_i}, \mu_i, \sigma_i^2) = \frac{D_i}{\sigma_i\sqrt{2\pi}(t-t_{0_i})}\exp\left(\frac{-[\ln(t-t_{0_i})-\mu_i]^2}{2\sigma_i^2}\right) \tag{1}$$

Additionally, each primitive is characterized by its starting and ending angle, θ_{s_i} and θ_{e_i}, and its angular variation can be described as:

$$\phi_i(t) = \theta_{s_i} + \frac{\theta_{e_i}-\theta_{s_i}}{2}\left[1+\text{erf}\left(\frac{\ln(t-t_{0_i})-\mu_i}{\sigma_i\sqrt{2}}\right)\right] \tag{2}$$

The trajectory that produces the handwritten trace $\vec{v}(t)$ is computed as the temporal overlap of each primitive's velocity $\vec{v}_i(t)$:

$$\vec{v}(t) = \sum_{i=1}^{N}\vec{v}_i(t) = \sum_{i=1}^{N}\begin{bmatrix}\cos\phi_i(t)\\ \sin\phi_i(t)\end{bmatrix}D_i\Lambda(t; t_{0_i}, \mu_i, \sigma_i^2) \tag{3}$$

To create our ground truth, each sample was reconstructed with the $\Sigma\Lambda$ model, and the lognormal parameters of each primitive were extracted, using the Sigma-Lognormal extractor proposed by [10]. Since each sample can be composed of a variable number of lognormal functions, we summarized the six parameters of each pen-down trajectory using four statistical measures: mean, standard deviation, maximum, and minimum.

C. Extraction of Deep Learning embeddings

1) Style-disentangled Transformer: SDT consists of three main components: a dual-head style encoder, a content encoder, and a Transformer decoder. The style and content encoders extract features from offline images, and the style decoder produces both character-wise and writer-wise style features through two separate heads. The Transformer decoder passes these features through a multi-head attention layer, whose outputs is fed to a Gaussian Mixture Model (GMM) that models the pen movement distribution autoregressively. Additionally, the multi-head attention layer updates at each iteration with the previous GMM output. We extracted the features at the final iteration of the GMM after it has passed through the multi-head attention layer.

2) Diffusion-Handwriting-Generation: DHG consists of two main components: a text-style encoder, and a diffusion probabilistic model. The text-style encoder, built on a pretrained MobileNetV2 model, extracts both visual and textual features from offline image representations. These extracted features are then fed into a diffusion U-Net model, which is composed of multiple downsampling blocks followed by upsampling blocks, using convolutional skip connections. We extracted features after the output of the last convolutional encoding layer of the diffusion model.

3) DeepWriteSYN: DWS is a Variational Autoencoder that uses Bidirectional Recurrent Neural Nets (RNNs) for both the encoder and decoder. Each RNN cell is built using HyperLSTM units, which not only predict the next token in a sequence but also dynamically generate the weights for the subsequent LSTM cell. We extracted the features from the bottleneck layer (mean and variance of the latent distribution) from the encoder output, after the bidirectional HyperLSTM layers.

D. Retrieval and Evaluation

Data retrieval was performed using KD-trees, an algorithm for efficient nearest-neighbor search. The data was split into 70% training and 30% testing. In total, four trees were created using the training data: one tree using the $\Sigma\Lambda$ ground-truth features, and three trees for the embeddings generated by SDT, DHG, and DWS. After construction, retrieval was performed by querying each tree with the test samples.

The retrieved examples from the embeddings of the three models were then compared against the retrieved examples from the $\Sigma\Lambda$ ground truth. To evaluate retrieval performance, we computed Precision@K, which measures how many relevant items are in the top K positions, and Rank-biased Overlap (RBO), which quantifies the similarity of two ranked lists, considering also their order.

[2]As it was done in the early development of the Kinematic Theory [15], we use "primitive" to refer to these decomposed units, as opposed to "stroke", to avoid confusion with the "stroke gestures" term which is very popular in HCI.

III. Experimental Results

Figure 1 shows a t-distributed stochastic neighbor embedding (t-SNE) projection of the ΣΛ ground-truth features of the $1-GDS dataset. Samples that are spatially close tend to exhibit similar velocity profiles, but they do not necessarily belong to the same symbol class. Although there is a noticeable separation between slow and fast samples, this distinction is not absolute. In some cases, slow samples are clustered closely with fast samples. This can occur when a complex symbol, such as a right curly brace drawn at high speed (as seen in sample 6), resembles a simpler shape, like a checkmark, when drawn slowly (as shown in sample 5).

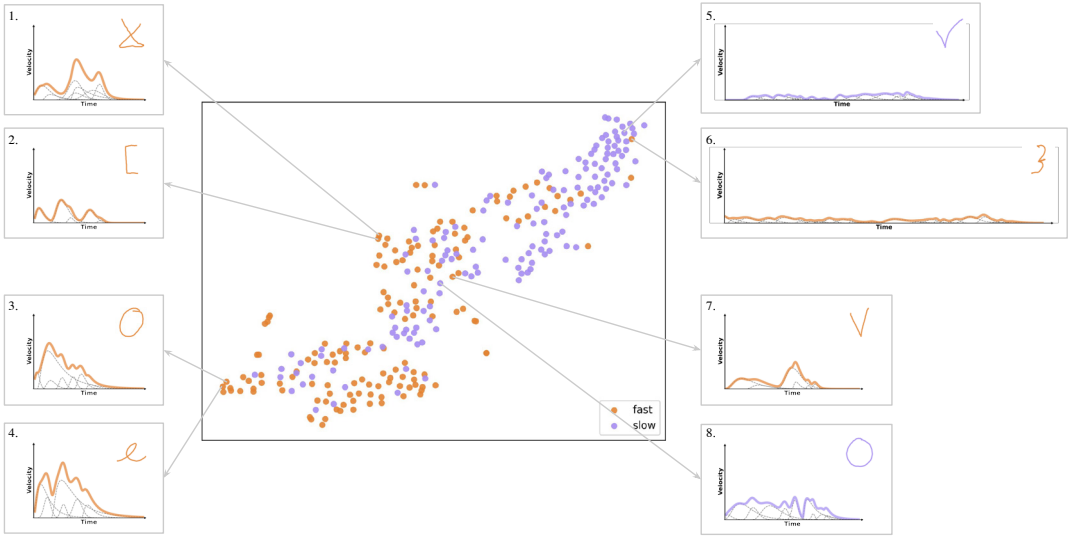


Fig. 1. t-SNE projection of the $1-GDS dataset using ΣΛ features as ground-truth. Selected samples are highlighted with their corresponding velocity profile. We can see that different gesture shapes may have similar articulation characteristics, as indicated by their velocity profiles, further showcasing the relevance of our work.

The retrieval experimental results are summarized in Table I. In addition to the results of the DL models, we consider two different baselines: *Random*, which represents the performance of a random retrieval system, and *Simple*, which retrieves samples based on one elementary feature, the number of points in each gesture.

TABLE I
RETRIEVAL PERFORMANCE RESULTS. TOP RESULT FOR EACH METRIC IS MARKED IN BOLD.

Model	Precision@5	Precision@10	Precision@20	RBO@5	RBO@10	RBO@20
SDT	0.049 ± 0.096	0.073 ± 0.083	0.145 ± 0.082	0.037 ± 0.087	0.049 ± 0.073	0.083 ± 0.062
DHG	0.056 ± 0.094	0.073 ± 0.083	0.118 ± 0.070	0.033 ± 0.073	0.049 ± 0.067	0.074 ± 0.059
DWS	**0.058 ± 0.100**	**0.084 ± 0.089**	**0.150 ± 0.089**	**0.045 ± 0.106**	**0.060 ± 0.086**	**0.092 ± 0.071**
Random baseline	0.015 ± 0.055	0.048 ± 0.066	0.095 ± 0.063	0.016 ± 0.057	0.026 ± 0.050	0.050 ± 0.047
Simple baseline	0.044 ± 0.088	0.077 ± 0.079	0.139 ± 0.082	0.029 ± 0.082	0.046 ± 0.062	0.077 ± 0.061

All DL models are able to outperform the Random baseline, by a considerable margin. However, all metrics have relatively low values across the three models, indicating weak relationship between DL embeddings and ΣΛ ground-truth data. It is worth noting that the standard deviations are high in all cases, suggesting there is considerable performance variability across different queries.

Among the models, DWS achieves the best performance in both retrieving relevant results and ranking them effectively. The performance differences between the three models are minor when retrieving 5 samples, but the gap increases slightly when retrieving 20 samples or more, especially when comparing DWS and DHG. Notably, the Simple baseline performs comparably to DL models, further reinforcing their weak relationship in terms of learned articulation features.

Figure 2 show examples of retrievals obtained with the ΣΛ ground-truth features, together with DWS, DHG, and SDT embeddings. The samples retrieved using the ground truth exhibit greater shape diversity, showcasing their focus on more than visual appearance. In contrast, the DL models appear to retrieve samples based on their visual shape, although this is less pronounced in DWS.

Figure 3 compares the velocity profiles of the queries with those of the retrieved results using the ground truth features. On average, the retrieved profiles visually align well with their corresponding queries.

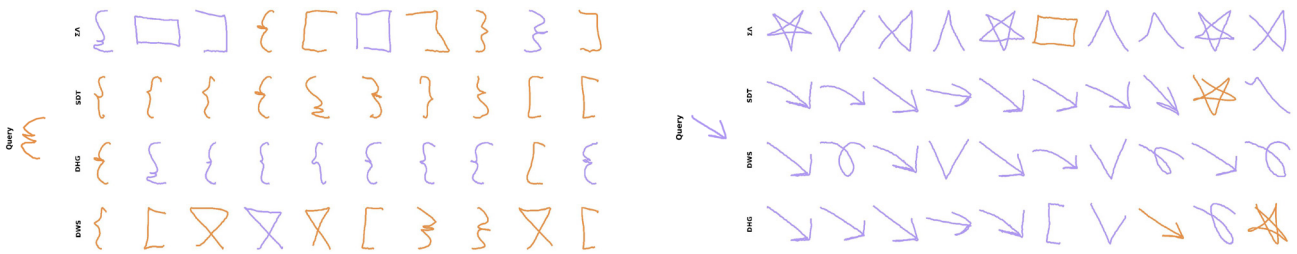


Fig. 2. Top-10 retrieval for two different queries, using ΣΛ, SDT, DHG and DWS features. Purple color indicates slow speed and orange indicates fast speed.

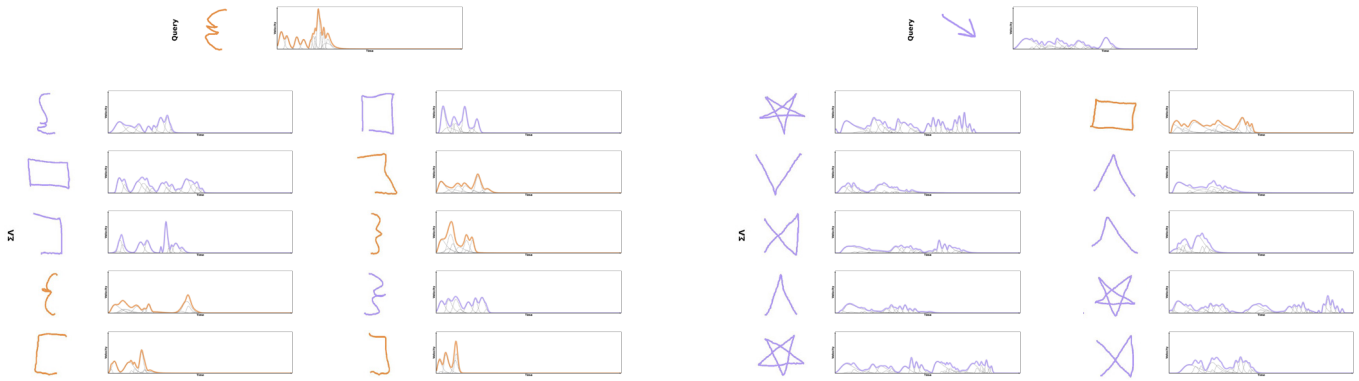


Fig. 3. Velocity profiles of the top-10 retrieval samples for two different queries, using ΣΛ ground truth features. Purple color indicates slow speed and orange indicates fast speed.

IV. Conclusion

In this work, we designed a retrieval task to investigate whether DL models for handwriting synthesis effectively capture articulation features. The results show a weak relationship between DL embeddings and ΣΛ features, suggesting that DL models do not adequately encode execution-related features. This could be due to them prioritizing other aspects of handwriting, like the spatial structure, but it could also be that they do encode articulation features but they differ from the ΣΛ features. One limitation is that we are comparing latent features which lie in entirely different feature spaces. A natural next step would be to use the DL models to fully reconstruct each sample, instead of extract the latent embeddings, then compute their corresponding ΣΛ parameters, and finally assess whether articulation features are preserved.

Acknowledgments

Research supported by Fonds National de la Recherche Luxembourg - FNR (SCRIPTOR project, grant AFR/22/17177001) and the European Innovation Council Pathfinder program (SYMBIOTIK project, grant 101071147).

References

[1] C. Appert and S. Zhai. Using strokes as command shortcuts: Cognitive benefits and toolkit support. *Conference on Human Factors in Computing Systems - Proceedings*, 2009.

[2] G. Dai, Y. Zhang, Q. Wang, Q. Du, Z. Yu, Z. Liu, and S. Huang. Disentangling writer and character styles for handwriting generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5977–5986, 2023.

[3] M. Djioua and R. Plamondon. Studying the variability of handwriting patterns using the kinematic theory. *Human movement science*, 28, 2009.

[4] F. Fang, H. Zhang, L. Zhan, S. Guo, M. Zhang, J. Lin, Y. Qin, and H. Fu. Handwriting velcro: Endowing ar glasses with personalized and posture-adaptive text input using flexible touch sensor. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(4), 2023.

[5] T. Flash and N. Hogan. The coordination of arm movements: an experimentally confirmed mathematical model. In *Journal of Neuroscience*, 1985.

[6] L. A. Leiva, D. Martín-Albo, and R. Plamondon. Gestures à go go: Authoring synthetic human-like stroke gestures using the kinematic theory of rapid movements. *ACM Trans. Intell. Syst. Technol.*, 7(2), 2015.

[7] L. A. Leiva, D. Martín-Albo, and R. Plamondon. The kinematic theory produces human-like stroke gestures. *Interacting with Computers*, 29(4), 2017.

[8] T. Luhman and E. Luhman. Diffusion models for handwriting generation, 2020.

[9] D. Meyer, J. Keith-Smith, S. Kornblum, R. Abrams, and C. Wright. Speed-accuracy tradeoffs in aimed movements: Toward a theory of rapid voluntary action. *Attention and Performance XIII*, 1990.

[10] C. O'Reilly and R. Plamondon. Development of a sigma-lognormal representation for on-line signatures. pattern recognit spec issue front handwrit recognit. *Pattern Recognition*, 42:3324–3337, 12 2009.

[11] R. Plamondon. A kinematic theory of rapid human movements. *Biol. Cybern.*, 72(4):309–320, 1995.
[12] R. Plamondon and M. Djioua. A multi-level representation paradigm for handwriting stroke generation. *Human Movement Science*, 25(4):586–607, 2006. Advances in Graphonomics: Studies on Fine Motor Control, Its Development and Disorders.
[13] A. Thomassen, G. van Galen, P. Keuss, and C. Grootveld. *Motor Aspects of Handwriting: Approaches to Movement in Graphic Behavior*. Acta psychologica. North-Holland, 1984.
[14] R. Tolosana, P. Delgado-Santos, A. Perez-Uribe, R. Vera-Rodriguez, J. Fierrez, and A. Morales. Deepwritesyn: On-line handwriting synthesis via deep short-term representations. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35:600–608, 2021.
[15] A. Woch and R. Plamondon. Characterization of bi-directional movement primitives and their agonist-antagonist synergy with the delta-lognormal model. *Motor control*, 14:1–25, 01 2010.
[16] L. Zhan, T. Xiong, H. Zhang, S. Guo, X. Chen, J. Gong, J. Lin, and Y. Qin. Toucheditor: Interaction design and evaluation of a flexible touchpad for text editing of head-mounted displays in speech-unfriendly environments. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 7(4), 2024.

Analyzing Angular Velocity Profiles of UR5e Robotic Arm Joints during Signing

Moises Diaz, Miguel A. Ferrer, José J. Quintana
iDeTIC, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain
{moises.diaz,miguelangel.ferrer,josejuan.quintana}@ulpgc.es

Adam Wolniakowski, Roman Trochimczuk, Konstantsin Miatliuk
Białystok University of Technology, Białystok, Poland
{a.wolniakowski,r.trochimczuk,k.miatliuk}@pb.edu.pl

***Abstract*— The Kinematic Theory of rapid movements postulates that the linear velocity of the end-effector after a rapid human movement can be modeled as a sum of lognormal functions. However, this assumption applies only to the end-effector's velocity, with no specification regarding the velocity of intermediate articulations. This raises the question of their speed profiles. In this paper, we investigate this by conducting experiments with a UR5e robotic arm, where the end-effector is programmed to mimic a human writer's signing trajectory and velocity. During execution, the robotic arm's controller records the angular velocities of its intermediate joint. The results show that the inverse kinematics solution for each joint produces an angular velocity profile as lognormal as that of the end-effector, in terms of SNR, SNR/NbLog and NbLog/sec. This result does not imply that human intermediate articulations move the same way but rather shows that such a solution is possible.**

Keywords*— *lognormality, angular velocities, joints, UR5e, robotic arm

I. Introduction

The observation of asymmetric bell-shaped velocity profiles in the end-effector after a rapid human movement has led to numerous studies and discussions aimed at formalizing the kinematic theory of rapid movements. This theory was developed in a series of four fundamental articles published between 1995 and 2003 by R. Plamondon [1]–[4]. It is based on two major hypotheses.

First, a given stroke produced by an end-effector is the output of a neuromuscular system consisting of a large number (n tending toward infinity) of coupled subsystems. Second, this coupling is mathematically represented by proportional cumulative time delays in command propagation, measured at various points along a given neuromuscular network: $T_j = (1 + \epsilon_j)T_{j-1}$.

If these two conditions are met, the Central Limit Theorem predicts that the impulse response of a neuromuscular system, observed in the velocity module of the end-effector, may be approximated by a lognormal function[1]: $v(t) = D\Lambda(t; t_0, \mu, \sigma^2)$, where the impulse response is described by:

$$\Lambda(t; t_0, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}(t - t_0)} \exp\left(\frac{-[\ln(t - t_0) - \mu]^2}{2\sigma^2}\right) \tag{1}$$

Here, t represents the time, D represents the amplitude of the impulsive stimulation command at time t_0, which initiates the central nervous system response. The parameters μ and σ correspond to the log-time delay and log-response time, respectively, reflecting the neuromuscular state of the individual.

The theory assumes that this primitive movement (or stroke) represents the impulse response of a neuromuscular system composed of a large number of coupled neural and muscular subsystems, modeling the behavior of an individual in perfect control of their movement.

This theory has been extensively validated in the velocity module of the end-effector for various graphonomics-based human rapid movements, including gestures, signatures, characters, handwriting, and even mouse movements. However, to the best of our knowledge, no experimental evidence has demonstrated some velocity patterns in the intermediate articulations responsible for executing the desired rapid movement in the end-effector.

To address this gap, we propose studying the angular velocity module of a robotic arm's joints when its end-effector executes a movement whose linear velocity module can be approximated by a sum of lognormal functions. First, we program the UR5e robotic arm to reproduce genuine handwritten signatures from the MCYT100 dataset [5], 2500 signatures in total. We then analyze and confirm that the robot-executed signatures can be effectively approximated by lognormal functions using the iDeLog method [6]. Once we verify that the end-effector velocity parameters adhere to the lognormality principle [7], we extract the angular velocities of each robotic joint.

As a chain of rigid links, the linear velocity of the joints is directly proportional to their angular velocity: $v_i = \omega_i \cdot R_i$, $i \in 1, \ldots, \ell$, where ω_i is the angular velocity of joint i, v_i its associated linear velocity, and $\ell = 5$, representing the five

[1]The full proof is detailed in [4].

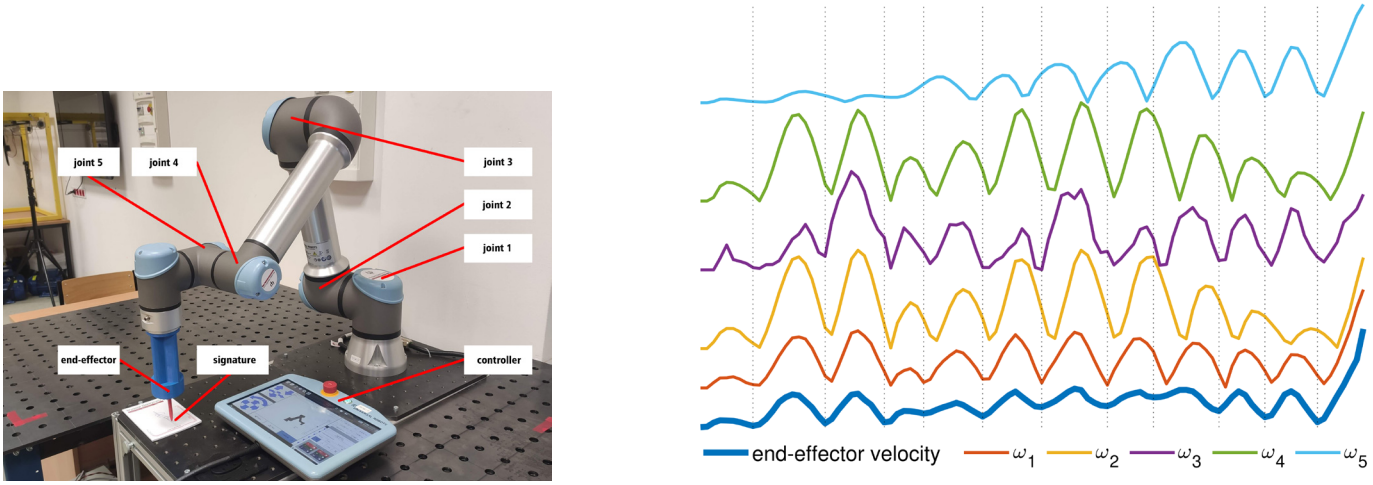


Fig. 1: View of the UR5e robot executing a signature (left) with numbered joints, and the corresponding end-effector linear velocity and joint angular velocities (right) recorded by the robot during signing.

degrees of freedom of the UR5e, studied in this work. Our experiments show that the angular velocity in the joints exhibits similar characteristics to the end-effector velocity in terms of signal-to-noise ratio, normalized by the number of lognormals, and the number of lognormals per second. The results suggest that when the linear velocity module of the end-effector follows a lognormal pattern, the joint velocities also adhere to the lognormality principle in a robotic arm. Notably, we do not impose specific joint movements; instead, they arise naturally from the geometric relationships of the robot's joints and defined after performing the inverse kinematics task of the UR5e manipulator.

II. Extracting Angular Velocities from the UR5e Robot During Signature Execution

Signatures in the MCYT100 [5] dataset were processed in CSV format, where each row contained tuples of the form $(t_i\,[\text{s}], x_i\,[\text{cm}], y_i\,[\text{cm}])$, representing the timestamp and the position of the pen tip in the X-Y plane. Here, $i = 1, \ldots, N$ corresponds to the row number, and N represents the total number of sampled data points. In this format, the z-coordinate of the pen tip is assumed to be zero. The trajectory waypoints were uniformly sampled in time at a frequency of $f_{sampling} = 100\,\text{Hz}$, as in the original dataset. To execute the writing movement using a robotic arm, this positional data needed to be transformed into a task frame of reference. Given the workspace of the UR5e robot arm, observed in Fig. 1 (left) we positioned the task frame at ${}^{\text{robot}}_{\text{task}}T = \text{Trans}(0, -0.5, 0)$.

The UR5e robotic arm is controlled using a custom ROS-based controller, named `kair_universalrobot`[2], which enables direct servo control of the joints in configuration space. This controller uses the UR5e primary and secondary interfaces to send URScript[3] commands to the controller and to read its current robot state during operation. The controller provides interfaces for both position and velocity control, as well as a trajectory action interface.

Since the controller operates in the joint space of the robot, the desired path is transformed into the configuration space of the UR5e robot. To achieve this, the inverse kinematics problem is solved numerically for each waypoint using the Jacobian approach.

UR5e is a relatively large and heavy robotic arm, not typically designed to executing swift and dexterous tasks such as signature writing. Signature movements impose a significant dynamic burden on the manipulator due to frequent and large acceleration spikes. To facilitate the execution of these movements on the UR5e robotic arm, we employed a reduced Jacobian approach, where only the translational part of the Jacobian matrix was used [8]. This method enforces the positional tracking, while allowing a certain degree of compliance in the orientation of the tool center point. Such compliance helps to distribute the efforts across the degrees of freedom of the robot. Additionally, we weighted the Jacobian columns to prioritize the use of the wrist joints, of they have lower inertia. The selected weights values were $\alpha = [0.01, 0.02, 0.05, 0.1, 0.2, 0.5]$.

The trajectory, expressed in joint space, is wrapped as a `FollowJointTrajectory` action and sent to the ROS controller, which handles interpolation and joint movement control using our custom interpolation library, `kair_trajectory`[4]. The controller employs linear interpolation with parabolic blend segments to ensure adherence to joint acceleration, effort, and velocity limits. It is important to emphasize that we do not directly control the angular velocity of the joints. Instead, the

[2] https://gitlab.com/dagothar/kair_universalrobot [Accessed: Apr. 29, 2025].
[3] https://www.universal-robots.com/products [Accessed: Apr. 29, 2025].
[4] https://gitlab.com/dagothar/kair trajectory [Accessed: Apr. 29, 2025].

inverse kinematics process determines the most suitable combination of angular and positional movements to reproduce the desired signature at the end-effector.

During trajectory execution, the controller reports and logs the robot's state. Among the recorded data are the angular velocities of each joint, denoted as $\omega = [\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6]$, in rad/s and defined in local joint space. Fig. 1 (left) displays a view of the UR5e robot, highlighting the joints whose angular velocities were recorded while executing a signature from the MCYT100 database. On the right side of the figure, the module of the linear velocity of the end-effector is shown, revealing a lognormal pattern as expected. The magnitude of the linear velocity of the end-effector is described in the global frame of reference, calculated as $v = ||^{\text{base}}\vec{v}_{\text{end-effector}}|| = ||J(\theta)\omega||$, in m/s, where $^{\text{base}}\vec{v}_{\text{end-effector}}$ is the linear velocity of the end effector wrt. the base of the robot, J is Jacobian matrix of the robot, and (θ, ω) are the angular positions and velocities of the robot joints.

Additionally, we present the angular velocity magnitudes of the other robot joints during the execution of the same signature. Interestingly, the velocity profiles of the joints also exhibit a lognormal pattern.

Comparing the magnitudes of joint angular velocities and the linear velocity of the end-effector may be inappropriate, as they are expressed in different spaces, joint space and Cartesian space, respectively, and have different units. Instead of comparing their values directly, in this work we compare the shapes of their magnitude profiles to analyze whether both exhibit lognormal characteristics, based on the decomposition of their peaks into lognormal components through iDeLog [6].

III. Experiments

A. Using iDeLog with unknown angular trajectories but known angular velocities

We used iDeLog [6] to extract the Sigma-Lognormal ($\Sigma\Lambda$) parameters from velocity data-linear velocity at the end-effector and angular velocity at the joints. However, this model requires a trajectory in the form of x-y signals as input, rather than velocity. To address this, iDeLog internally derives the velocity from the given trajectory, which is then modeled using lognormal functions.

In our case, we directly recorded the angular velocities at the joints but not the corresponding trajectories. To avoid modifying the iDeLog code, we generated a fictitious trajectory for each joint that reproduces the recorded velocity.

Let ω_i be the magnitude of the angular velocity at joint i. We constructed a circular trajectory whose derivative reproduces ω_i as follows:

1) The local maxima (peaks) of ω_i were detected, and the positions p_k between peaks were stored to compute the radius of the circumference as $r = 10 \cdot \bar{p}$, where $\bar{p}$ is the average position difference between consecutive local maxima, and 10 is a scaling factor.
2) The angular positions q_i were estimated as the cumulative sum of the angles, computed as $q_i = \text{cumsum}\,[0; \omega_i/r]$.
3) A fictitious circular trajectory was then developed as $(x_f, y_f) = r \cdot (\cos(q_i), \sin(q_i))$.

This approach enables the use of iDeLog when the trajectory is unknown (as in our case) and only the velocity is available.

B. Adherence to lognormality in the end-effector velocity module of the Robot

First, we confirm that the linear velocity of the robot's end-effector follows the lognormality principle. Although previous studies have shown that the linear velocity of handwriting in the MCYT-100 dataset exhibits lognormal patterns, it is essential to ensure that the linear velocity of the robot's end-effector, which was programmed to reproduce the signatures, can also be approximated using the $\Sigma\Lambda$ model. In practice, certain metrics are used to validate lognormality. In this study, we employ three metrics:

1) The signal-to-noise ratio, SNR, which compares the observed and estimated velocity magnitude. A higher value indicates greater adherence to the kinematic theory, with a practical validation threshold of approximately 15 dB.

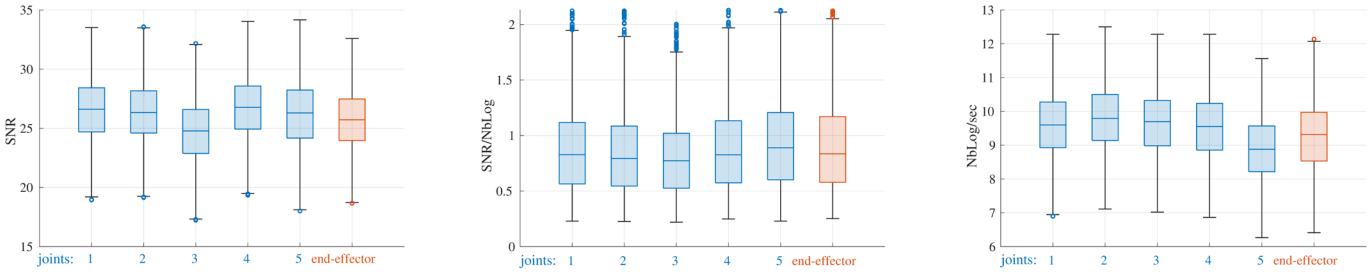


Fig. 2: Experimental results based on $\Sigma\Lambda$ parameters using iDelog: The blue box plots represent the angular velocity of each robot joint, while the red box plots show the linear velocity of the robot's end-effector.

2) The SNR/NbLog ratio, which quantifies movement fluidity, where higher values indicate smoother motion.
3) The NbLog/sec ratio, which represents the number of strokes per second in the signatures. Normalizing NbLog by signature duration ensures that its analysis is independent of signature length.

Fig. 2 shows the distribution of these three parameters (in red). The first quartile of SNR is above 23 dB, confirming a strong match between the end-effector velocity magnitude and a purely lognormal velocity pattern. The first and third quartiles for SNR/NbLog are 0.60 dB and 1.30 dB, respectively, characterizing the movement fluidity in our dataset. Finally, the number of lognormal strokes per second were found between $8.50\,s^{-1}$ and $9.98\,s^{-1}$, based on the first and third quartiles.

As expected, the results show that the linear velocity of the end-effector follows a lognormal profile, aligning closely with the $\Sigma\Lambda$ model.

C. Lognormal analysis of angular velocities in the joints

As shown in Fig. 1 (right), the angular velocity profile in the five joints exhibits a bell-shaped curve, characteristic of lognormal movements. To investigate this further, the data was analyzed using iDeLog to assess its lognormality. The results are presented in Fig. 2, in blue. As seen, the values for the end-effector are similar to those obtained for the angular velocities of the joints.

For SNR, to define the range of results across the five joints, the lowest first quartile was 22.88 dB (observed in the 3rd joint), while the highest third quartile was 28.57 dB (found in the 4th joint).The interquartile range (IQR) remains consistent at 3.80 ± 0.05 dB across all joints.

Regarding SNR/NbLog, the five joints show very similar results, with a stable IQR of 0.67 ± 0.05 dB. The lowest first quartile (0.53 dB) and highest third quartile (1.21 dB) were found in the 3rd and 5th joints, respectively. These findings indicate that movement fluidity in the joints is similar to that of the end-effector.

Finally, for NbLog normalized per second, the IQR was $1.37\pm0.01\,s^{-1}$, with the lowest first quartile ($8.20\,s^{-1}$) in the 5th joint and the highest third quartile ($10.50\,s^{-1}$) in the 2nd joint.

In summary, no significant differences were found between these metrics when comparing the end-effector velocity module with the angular velocity modules of the joints.

IV. Conclusion

The study of the end-effector velocity profile is valuable for understanding human movement and provides useful information in various fields, such as e-health, e-security, and e-learning. Additionally, analyzing the velocity profiles of intermediate joints can provide complementary insights beyond those obtained from the end-effector alone.

In this study, we took an initial step toward addressing this challenge by using a UR5e robotic arm programmed to replicate natural handwriting with five degrees of freedom. The robotic tip mimicked human handwriting with a pencil, both in trajectory and velocity. The movement of the intermediate joints was determined through inverse kinematics.

Interestingly, our results showed that the angular velocities of the joints followed a lognormal distribution, similar to the end-effector velocity when analyzed through SNR, SNR/Nblog, and Nblog/sec. This suggests that a feasible solution for the lognormal motion of the end-effector could be that all joints exhibit lognormal angular velocities. However, this does not necessarily align with human movement strategies.

As a next step, we plan to experimentally analyze human joint velocities using motion capture (MOCAP) sensors, particularly during rapid graphonomic movements of the hand (end-effector). This will provide deeper insights into how humans coordinate articulation movements in relation to end-effector motion.

Acknowledgment

This research was partly supported by the PID2023-146620OB-I00, funded by MICIU/AEI 10.13039/501100011033 and the European Union's FEDER program, and partly by the CajaCanaria and la Caixa (2023DIG05).

References

[1] Réjean Plamondon. A kinematic theory of rapid human movements: Part I. movement representation and generation. *Biological cybernetics*, 72:295–307, 1995.
[2] Réjean Plamondon. A kinematic theory of rapid human movements: Part II. movement representation and generation. *Biological cybernetics*, 72:309–320, 1995.
[3] Rejean Plamondon. A kinematic theory of rapid human movements: Part III. kinetic outcomes. *Biological Cybernetics*, 78(2):133–145, 1998.
[4] Rejean Plamondon, Chunhua Feng, and Anna Woch. A kinematic theory of rapid human movement. Part IV: a formal mathematical proof and new insights. *Biological cybernetics*, 89:126–138, 2003.
[5] Javier Ortega-Garcia, J Fierrez-Aguilar, D Simon, J Gonzalez, Marcos Faundez-Zanuy, V Espinosa, A Satue, I Hernaez, J-J Igarza, C Vivaracho, et al. MCYT baseline corpus: a bimodal biometric database. *IEE Proceedings-Vision, Image and Signal Processing*, 150(6):395–401, 2003.
[6] Miguel A Ferrer, Moises Diaz, Cristina Carmona-Duarte, and Rejean Plamondon. iDeLog: iterative dual spatial and kinematic extraction of sigma-lognormal parameters. *IEEE transactions on pattern analysis and machine intelligence*, 42(1):114–125, 2018.
[7] Réjean Plamondon. The lognormality principle: a personalized survey. In *The Lognormality Principle and Its Applications in e-security, e-learning and e-health*, pages 1–39. World Scientific, 2021.
[8] Bruno Siciliano. Kinematic control of redundant robot manipulators: A tutorial. *Journal of intelligent and robotic systems*, 3:201–212, 1990.

Human Movement as a State Transition

Moussa Djioua

Département de Génie Électrique,Polytechnique Montréal
2500 Chemin de Polytechnique, Montréal,H3T 1J4,Québec, Canada
m.djioua@polymtl.ca

***Abstract*— Rapid human movement has been extensively studied through its stereotypical bell-shaped velocity profile. U sing t he central limit theorem and the agonist-antagonist synergy paradigm, it has been shown that the Delta-Lognormal model represents its convergence limit, while the Sigma-Lognormal model extends this framework to analyze, in velocity space, trajectories of composite movements such as handwriting, signatures, and psychophysical experiments. This paper aims to analyze human movements of an end-effector within the framework of state transition theory. The end-effector is viewed as a mobile point, whose trajectory corresponds to the step response of a neuromuscular system controlled by the central nervous system. The neuromusculoskeletal nature of the motor system exhibits four distinct dilations in response to a step command, which can be observed through its temporal shift, sigmoid profile, a nd c urvature i n p lanar s pace. In the case of ballistic movements, from the departure to the target point, the recorded planar (2D) and spatial (3D) stereotypical patterns arise as projections of a sigmoidal joint rotation profile b etween t wo r esting p ositions. C onsequently, h andwriting, s ignatures, a nd o ther composite trajectories, produced by the end-effector, result from a controlled sequence of state transitions. The Sinerf model is proposed to describe any planar trajectory of human movement using a complex function and to model these trajectories in 3D space using a quaternion function.**

Keywords*— *Kinematic Theory of Human Movement, Sigma-Lognormal model, Sinerf model, state transition theory.

I. Introduction

The ballistic movement has been extensively studied, and several models have been proposed to describe its bell-shaped velocity profile. By considering the neuromusculoskeletal structure of the motor system, it has been demonstrated that the Delta-Lognormal function represents the causal limit profile toward which all other models converge [3]–[7]. Consequently, the Sigma-Lognormal model was proposed to analyze complex planar movements, such as handwriting and signatures, in velocity space [8], [9], [12]. In psychophysics, rapid movements recorded in pointing experiments are considered a baseline for analyzing simple human movements, while the analysis of handwriting and signature movements is first performed through the examination of oscillatory movements at maximum velocity. By understanding the nature of motor control through these two analyses, the movement trajectory is viewed as resulting from the control of a sequence of rapid movements. This paper presents a reformulation of human movement modeling by describing its trajectory in the space where it was recorded. Thus, the trajectory of a planar movement is modeled by a complex function $\mathbf{z}(t)$, while that of a spatial movement is represented by a quaternion function $\mathbf{q}(t)$.

II. Movement as a state transition

The intention to move the end-effector of a limb between two positions P_0 and P_1, separated by a distance A, is viewed as a state transition of a mobile point between two resting states. This intention is modeled by $P_1 = P_0 + Au(t)$ where $u(t)$ represents the step function (Heaviside function) that defines the temporal transition. However, due to the neuromusculoskeletal nature of the motor system, this state transition undergoes four dilations and triggers an antagonistic reaction proportional to the intensity of the change. The first two dilations are purely neural. As illustrated in Fig.1-a, they are observed through the delay t_0 between the initiation of the transition and the moment the command is sent to the motor areas, as well as the propagation delay $t_1 - t_0$, which separates the moment when the control system sends the command and the mechanical onset of movement. The third dilation reflects the inertia of biological systems, which oppose instantaneous changes and behave as low-pass filter. Thus, under the assumption that the motor system is not in an oscillatory mode, the step command is transformed into a sigmoid function, which is best observed in velocity space as a generally asymmetric bell-shaped profile. The fourth dilation refers to the curvature of the observed trajectory, expressing a temporal variation of direction. Indeed, the mobile point acts as the end-effector of a multi-articulated skeletal system. Its trajectory results from the combined rotations of multiple joints, which, over a short period, can be approximated as the rotation of a single equivalent joint, designated by a virtual articulation. It is assumed that the segment length, connecting the joint to the end-effector, remains constant during a ballistic movement. Consequently, the resulting trajectory describes an arc of a circle, whose curvature C is normalized by the total angular rotation of the articulation $\Theta = CA$. Finally, according to the kinematic theory of human movement [3], [4], an antagonistic reaction slows down the final phase of the movement. Under certain conditions, this reaction can induce transient or sustained oscillatory regimes, such as tremors.

III. Limit profile of the sigmoid

The kinematic theory of movement predicts that the limit velocity profile of a rapid movement follows a lognormal function [3]–[6]. However, when the pointing movement is not fast, the velocity profile becomes multimodal. Since motor units —arranged in parallel— generate the forces required to move the skeletal system, the overall profile of temporal dilation depends on the level of synchrony among these units. Each motor unit is assumed to produce an infinitesimal displacement, such that the total displacement associated with the state transition results from their grouped actions. The resulting temporal profile emerges from their degree of synchrony: perfect synchronization produces a modal profile, whereas the presence of a multimodal profile suggests either (1) movement learning period, (2) voluntary scheduling of motor unit activations (following the speed-accuracy tradeoff principle) or (3) dysfunction of certain motor units due to fatigue or other underlying factors. Thus, the control of state transitions is achieved through the synchronization of motor units, which directly influences the shape of the sigmoid function —most clearly observed in the velocity profile. By definition, the primitive of the sigmoid function describing temporal dilation is given by:

$$s(t-t_0,\tau,\sigma)=\frac{1}{2}\left(1+\operatorname{erf}\left[\frac{\log\frac{t-t_0}{\tau}}{\sigma\sqrt{2}}\right]\right) \tag{1}$$

Let t_0 represent the neural reaction delay relative to the reference time, and τ the geometric mean or arithmetic median, indicating the moment when half amplitude of the state transition has been reached. The time $\tau e^{-3\sigma}$ corresponds to the second purely neural dilation of the peripheral system, while $\tau e^{3\sigma}$ represents the total duration of neuromuscular dilation, and $2\tau\sinh 3\sigma$ denotes the purely mechanical dilation of the transition. The parameter σ characterizes the degree of asymmetry. Specifically, if the velocity profile of the state transition were symmetric, the mechanical duration would be equal to 2τ. The quantity $\sinh 3\sigma$ extends this duration, introducing a positive asymmetry. The general form of the dilation function is modeled as follows:

$$s(t;p_m)=\sum_{i=1}^{m}s(t-t_{0_i},\tau_i,\sigma_i)-s(t-t_{0_-},\tau_-,\sigma_-) \tag{2}$$

Let $p_m=\{\{t_{0_i},\tau_i,\sigma_i\}_{i=1\ldots m},\{t_{0_-},\tau_-,\sigma_-\}\}$ the parameter vector grouping the m triplets of agonist parameters and the triplet of antagonist parameters of the first three temporal dilations.

IV. Case of Rotational Movements

In this kind of movements, the recorded trajectory is a one-dimensional signal, which can represent, for example, the rotation of a specific articulation, such as the wrist. A rotational movement corresponds to a state transition of amplitude Θ between two angular positions, θ_0 and θ_1, of the articulation. The control signal governing this transition is defined by : $\theta_1=\theta_0+\Theta u(t)$. This control signal undergoes three temporal dilations, and the recorded trajectory at the end-effector is given by:

$$\theta(t)=\theta_0+\Theta s(t;p_m)\rightarrow\theta_1\ ,t\rightarrow\infty \tag{3}$$

The Fig.1-a, illustrates the temporal dilations undergone by the step command in the one-dimensional space, and Fig.1-c a typical fitting result of a wrist joint rotation by this model.

V. Case of planar movements

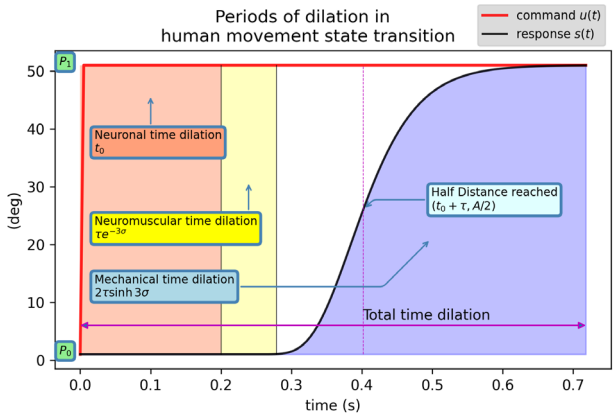
Consider the simple case of an articulation rotation, occurred in a plan and described by Equ.4. The planar trajectory of its terminal effector follows an arc of a circle, transitioning from point P_0, to point P_1. In the absence of dilations, this trajectory is given by:

$$\mathbf{z}_1=\mathbf{z}_0+Ae^{\jmath\theta_0}u(t) \tag{4}$$

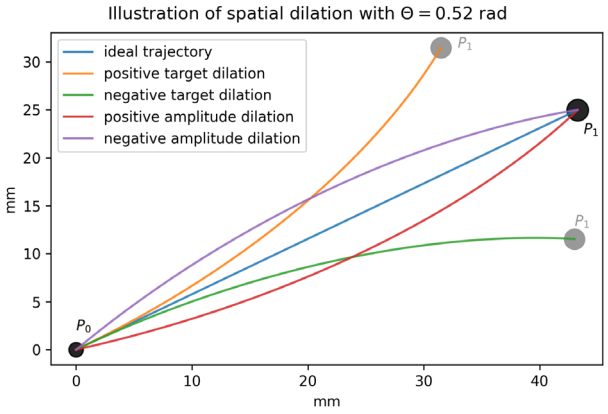
where A represents the distance, and θ_0 denotes the directional angle of the trajectory between the resting points P_0(position $\mathbf{z}_0$) and P_1 (position $\mathbf{z}_1$). However, this planar movement undergoes temporal dilations, where $u(t)\rightarrow s(t;p_m)$, and a spatial dilation, which is a direct consequence of the multi-joint structure of the motor system. This spatial dilation manifests as curvature $\Theta\in[-2\pi,2\pi]$, measured in radians, with its kernel $\operatorname{expc}(\Theta)$—referred to as the exponential cardinal—being a Hilbert analytical signal with its transformation $\mathcal{H}$, expressed as:

$$K_{\Theta_t}=\operatorname{expc}(\jmath\Theta_t)=\operatorname{sinc}(\Theta_t)+\jmath\mathcal{H}\{\operatorname{sinc}(\Theta_t)\}=\frac{e^{\jmath\Theta_t}-1}{\jmath\Theta_t}=e^{\jmath\Theta_t/2}\operatorname{sinc}(\Theta_t/2) \tag{5}$$

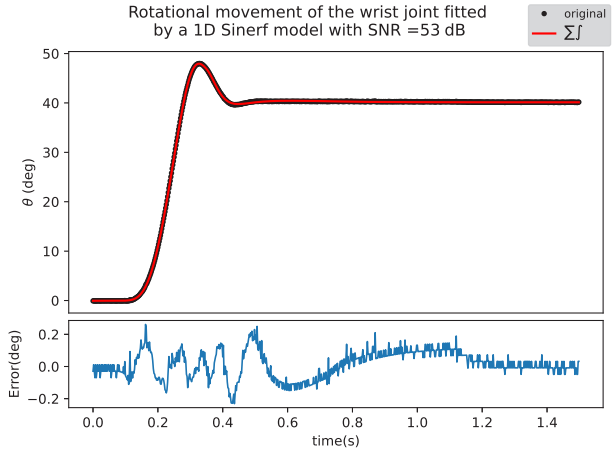
Spatial dilation can be interpreted in two distinct ways: (1) as described by Eq.5, the amplitude A remains constant while the target position P_1 is displaced from its intended location; and (2) the position P_1 remains fixed, but the amplitude A is scaled by a factor $\alpha=\frac{e^{-\jmath\Theta/2}}{\operatorname{sinc}\Theta/2}$. In this case, the kernel of the spatial dilation becomes $K_{\Theta_t}=e^{\jmath\frac{\Theta}{2}(s(t)-1)}\frac{\operatorname{sinc}(\Theta_t/2)}{\operatorname{sinc}(\Theta/2)}$.



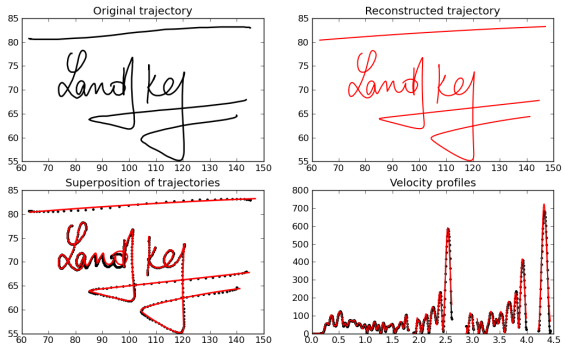
(a)



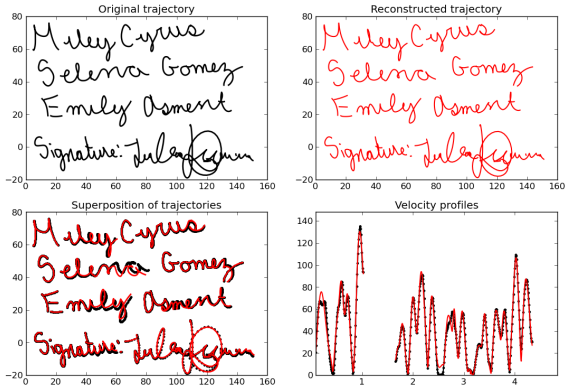
(b)



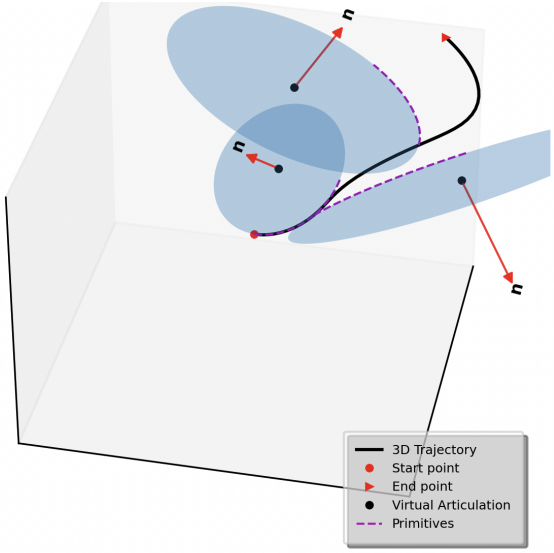
(c)



(d)



(e)



(f)

Fig. 1. (a) Illustration of the different temporal dilations affecting the command during a state transition (b) Illustrations of spatial dilation affecting trajectory curvature (c) Typical fitting result of angular trajectory with 1D version of a Sinerf model, performed by a rapid wrist movement (d) Typical fitting result of a signature decomposed using the 2D version of the Sinerf model, reprint from [11] (e) Typical preteen's handwriting sample fitted by a 2D version of Sinerf model , reprint from [11] (f) Illustration of a typical synthesis of a three-dimensional trajectory using the 3D version of the Sinerf model.

Fig.1-b illustrates the two versions of spatial dilation. The model presented here uses the kernel defined in Eq.5.

The presence of a virtual articulation, resulting from the postural combinations of skeletal joints, influences the movement direction, which follows the same sigmoidal profile as the amplitude. Thus, Eq.6 highlights the simultaneous effect of curvature on the temporal profiles of both direction and amplitude.

The total trajectory of the state transition is given by:

$$\mathbf{z}(t) = \mathbf{z}_0 + Ae^{\jmath\theta_0} s(t) K_{\Theta s(t)} = \mathbf{z}_0 + As(t)\,\mathrm{sinc}\left(\frac{\Theta s(t)}{2}\right) e^{\jmath\left(\theta_0 + \frac{\Theta}{2}s(t)\right)} \rightarrow \mathbf{z}_1 \;\; , t \rightarrow \infty \tag{6}$$

In the absence of dilations $s(t) = u(t)$ and $\Theta = 0 \rightarrow K_0 = 1$, this reduces to the original command $\mathbf{z}(t) = \mathbf{z}_0 + Ae^{\jmath\theta_0} u(t)$

A. Generalized Case of Planar Motion

Let (t, x, y) be the trajectory of any arbitrary movement executed on a plane, such as on a digitizing tablet, given by $\mathbf{z}(t) = x(t) + \jmath y(t)$. It is modeled as a sequence of N planar state transitions, expressed as:

$$\mathbf{z}(t) = \mathbf{z}_0 + \sum_{i=1}^{N} A_i e^{\jmath\theta_{0_i}} s_i(t, p_m) K_{\Theta_i s_i(t, p_m)} \tag{7}$$

This model demonstrates that the x and y coordinates are simply the projections of a sigmoidal rotation $\theta(t)$ of the articulation —whether real or virtual—occurring between two equilibrium positions. Given that the chosen sigmoidal function represents the limit profile, according to kinematic movement theory, this model is named Sinerf, where: "Sin" refers to the geometric projection, "erf" (error function) which represents the shape profile of the state transition [11]. Sinerf is an extension of the Sigma-Lognormal model [8]–[10]. Unlike this latter, which describes kinematics in velocity space, the Sinerf model operates directly in the original recorded space. In fact, the derivation or integration of original data should serve only for visual representations or as preprocessing steps for initial solution estimation in inverse problem-solving. Modeling the data within its original recording space preserves the intrinsic structure of the trajectories. Fig.1-d,e provide an overview of the handwriting and signature fitting using the planar version of the Sinerf model [11].

B. Comparing Sinerf and Sigma-Lognormal models

Let us now revisit the comparison of the two models in greater detail. Let us consider the case of a single pen-stroke trajectory, expressed by the complex function $\mathbf{z}(t)$. For such a case $s(t) > 0$, we reformulate Eq.6 as follows:

$$\mathbf{z}(t) = \mathbf{z}_0 + Ae^{\jmath\theta_0} s(t) \frac{e^{\jmath\Theta s(t)} - 1}{\jmath\Theta s(t)} = \mathbf{z}_0 + \jmath\frac{A}{\Theta} e^{\jmath\theta_0} \left(1 - e^{\jmath\Theta s(t)}\right) \tag{8}$$

Assuming that $s(t)$ contains only one sigmoid component, and using the expression from Eq.3, representing the joint rotation angle, the derivative of $\mathbf{z}(t)$ yields $\frac{d\mathbf{z}(t)}{dt} = \dot{\mathbf{z}} = A\Lambda(t)e^{\jmath(\theta(t))}$. For a trajectory composed of N pen-strokes, the velocity profile of the trajectory is expressed as

$$\dot{\mathbf{z}}(t) = \sum_{i=1}^{N} A_i \Lambda_i(t) e^{\jmath(\theta_i(t))} = \sum_{i=1}^{N} A_i \Lambda_i(t) \cos\theta_i(t) + \jmath \sum_{i=1}^{N} A_i \Lambda_i(t) \sin\theta_i(t) \tag{9}$$

Eq.9 expresses the tangential velocity vector of the trajectory $\mathbf{z}(t)$. It provides an alternative formulation to Equations (1) through (8) of the Sigma-Lognormal model described in [12]. The key differences between the Sinerf and the Sigma-Lognormal models lie in the following: (1) in the latter, each stroke contains only one component — the antagonistic component is treated as a separate pen-stroke; (2) the trajectory is represented in the velocity space using the magnitude of the velocity, derived through numerical differentiation and filtering; and (3) the model reconstructs the trajectory via numerical integration. These operations — filtering, differentiation, and integration — may degrade certain kinematic information embedded in the trajectory, which could reveal subtle variations in fine motor control.

VI. Case of Three-Dimensional Movements

Due to the multi-jointed structure of the limbs, the primitive of a rapid movement trajectory is assumed to only occur within a single plane. Indeed, in biomechanical point of view, skeletal joint undergoes rotational motion within a plane determined by the adjacent bone segments and follow preferred directions. However, the spatial configuration and activation sequences of motor units enable, through vectorial superposition, the execution of movements across a wide range of oriented planes. Furthermore, this mechanism facilitates the formation of virtual joints, arising from the combined rotations of anatomical joints. Thus, the 3D trajectory of the end-effector is a concatenation of planar movements, each one occurring within a scheduling sequence of oriented planes. By definition, the trajectory of a 3D rapid movement is mathematically represented as a complex function $\mathbf{z}(t)$ within an oriented plane,characterized by its unit normal vector $\mathbf{n}$.

A. Representation by complex functions

Let $\mathbf{x}_t = (t, x, y, z)$ be the vector of points defining a 3D trajectory within an oriented plane characterized by its normal vector $\mathbf{n}$. We define two unit vectors, $(\mathbf{u}, \mathbf{u}_\perp)$, such that, together with $\mathbf{n}$, they form a 3D Euclidean reference frame. Let $\mathbf{z}_t$ represents the trajectory within the plane, as described by Eq.7. Consequently, the 3D trajectory can be expressed as a function of the complex variable $\mathbf{z}_t$ in the following form:

$$\mathbf{x}_t = \Re\{\mathbf{z}_t\}\mathbf{u} + \Im\{\mathbf{z}_t\}\mathbf{u}_\perp \tag{10}$$

B. Representation by quaternions

Let q_t be the quaternion that describes the trajectory $\mathbf{x}(t)$ of the ballistic movement, and let q_t^* be its conjugate, such that:

$$q_t = q_0 + \mathbf{q}_t = \cos\frac{\Theta_t}{2} + \mathbf{n}\sin\frac{\Theta_t}{2} \tag{11}$$

Let $\mathbf{x}_0$ be the initial position with respect to the Euclidean reference frame, $\mathbf{z}_0$ the initial position relative to the oriented plane defined by the unit basis vectors $(\mathbf{u}, \mathbf{u}_\perp)$, and $\hat{\mathbf{x}}_0$ the initial position with respect to the virtual articulation at position $\mathbf{x}_{va} = \mathbf{x}_0 - \hat{\mathbf{x}}_0$, such that $\hat{\mathbf{x}}_0 = \Re\{\mathbf{z}_0\}\mathbf{u} + \Im\{\mathbf{z}_0\}\mathbf{u}_\perp$ Then, the rotation around the normal vector $\mathbf{n}$ displaces the end-effector from its initial position $\mathbf{x}_0$ to the final position $\mathbf{x}_1$, given by:

$$\mathbf{x}_t = \mathbf{x}_0 + (q_t\hat{\mathbf{x}}_0 q_t^* - \hat{\mathbf{x}}_0) \rightarrow \mathbf{x}_1 \quad , t \rightarrow \infty \tag{12}$$

To generalize, the total trajectory of a sequence of state transitions is expressed as a succession of rotations around the unit normal vectors $\mathbf{n}_i$ of N successive oriented planes (see Fig.1-f). This trajectory is modeled as:

$$\mathbf{x}_t = \mathbf{x}_0 + \sum_{i=1}^{N} \left(q_{t_i}\hat{\mathbf{x}}_{0_i} q_{t_i}^* - \hat{\mathbf{x}}_{0_i}\right) \tag{13}$$

VII. Conclusion

The human movement in question is the one resulting from the intention to move an end-effector between two equilibrium positions. This displacement, triggered by a step command, is viewed as a state transition of the motor system. As a biological system, it inherently resists to instantaneous changes. Its free response is observed in its kinematics through a sequence of three temporal and one spatial dilations. In movements triggered by external stimuli (audio, visual , tactile, etc.), the neural processes underlying perception and decision-making can be inferred from the movement trajectory through kinematic analysis via a time delay t_0, which represents the duration between the appearance of the stimulus and the initiation of movement. Additionally, the parallel organization of motor units in the neuromuscular system shapes the mechanical state transition of the end-effector. This transition is reflected in its sigmoid profile, resulting from the temporal dilation of the step command. It provides insight into motor control by revealing the synchronization of motor unit activity. The observed mono-modal and multi-modal velocity profiles serve as measures of this synchrony. Furthermore, the multi-jointed structure of the skeletal system causes the trajectory of human movement to be a projection onto Euclidean axes of an underlying sigmoid-shaped articulation rotation. The proposed Sinerf model was developed to extend the Sigma-Lognormal model, transferring it from velocity space to the original position space, where the trajectory is represented as a concatenation of pen-strokes. This model was specifically designed to meet the needs of 3D trajectory modeling in anthropomorphic robotics, by taking human biomechanics into account. Thus, the control of movement's direction, amplitude, and velocity profiles is achieved through the synchronization activities of redundant actuators. Although, originally designed for trajectory synthesis in anthropomorphic field, it can be used to efficiently analysis rapid and slow human movements by considering each them as executed with a multi-modal velocity profile. This multi-modality occurs through concatenation on a same temporal scale of lognormale primitives.

References

[1] Abend, W., Bizzi, E. and Morasso, P. "Human arm trajectory formation", Brain 105, 331–348, 1982.
[2] Morasso, P. "Spatial control of arm movements", Experimental Brain Research 42, 223–227 1981.
[3] Plamondon,R."A kinematic theory of rapid human movements:Part I: Movement representation and generation". Biological Cybernetics.72:295-307,1995.
[4] Plamondon, R. "A kinematic theory of rapid human movements:Part II. Movement time and control". Biological Cybernetics. 72:309-320, 1995.
[5] Plamondon, R. "A kinematic theory of rapid human movements:Part III. Kinetic outcomes". Biological Cybernetics..78:133-145, 1998.
[6] Plamondon,R.,et al."A kinematic theory of rapid human movement.Part IV:a formal mathematical proof and new insights".Biol.Cybern.89,126–138 2003.
[7] Djioua, M. Plamondon R "The limit profile of a rapid movement velocity", Human Movement Science (2009).
[8] Djioua M. , O'Reilly C. and Plamondon R. "An interactive trajectory synthesizer to study outlier patterns in handwriting recognition and signature verification," Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), vol. 1, pp. 1124- 1127. 2006.
[9] O'Reilly, C. , Plamondon, R. "Development of a sigma-lognormal representation for on-line signatures". Pattern Recognit Spec Issue Front Handwrit Recognit. Pattern Recognition. 42. 3324-3337. 10.1016 .2008.
[10] Djioua M. Modélisation, Analyse et Synthèse de la Cinématique du Mouvement Humain. Editions universitaires europeennes, 2010.
[11] Djioua M. "Modèle Sinerf: théorie et performances", unpublished, 2011.
[12] O'Reilly C., Plamondon R., and Faci, N., "The Lognometer : A New Normalized and Computerized Device for Assessing the Neurodevelopment of Fine Motor Control in Children," 2022 26th International Conference on Pattern Recognition (ICPR), Montreal, QC, Canada, 2022, pp. 952-958, doi: 10.1109/ICPR56361.2022.9956357.





Session

Forensic & Security

Analysis of Inclinations in Genuine Signatures and Skilled Forgeries

Marcos Faundez-Zanuy Tecnocampus, Universitat Pompeu Fabra
faundez@tecnocampus.cat

Abstract- This paper investigates the inclination characteristics of genuine and skilled forgery online signatures, focusing on their implications for biometric recognition systems. Our analysis reveals that while inclination removal through normalization does not significantly impact recognition accuracy for identification and verification tasks, it highlights notable differences between genuine signatures and skilled forgeries. Skilled forgeries exhibit significantly greater variability in inclination angles compared to genuine signatures.

Keywords- ***online signature skilled forgeries inclination.***

I. METHODS

Inclination estimation can play a crucial role in online signature analysis, particularly in the context of biometric recognition [1]. The inclination of a signature, which reflects the overall tilt or inclination axis of the handwriting, is a key characteristic that can be influenced by a user's writing style, habits, and ergonomics. Accurately estimating this inclination allows for normalization through a process known as de-rotation, where the signature is adjusted to align with a standard orientation. This normalization helps reduce intra-user variability, making genuine signatures more consistent and, therefore, easier to match during verification processes. However, while de-rotation can enhance the robustness of biometric systems by mitigating human variability, it may also inadvertently remove or obscure relevant information. The inclination of a signature is often a distinctive and stable trait for an individual, contributing to the distinctiveness of their handwriting. By eliminating this characteristic, important discriminatory information could be lost, potentially impacting the ability to differentiate between genuine signatures and forgeries. This trade-off highlights the importance of studying the relevance of signature inclination in biometric systems. Understanding how inclination contributes to the individuality and stability of a signature can help balance the benefits of normalization with the preservation of meaningful biometric features. In this paper, we analyze the variability of inclinations in genuine signatures and skilled forgeries, as well as the impact of removing this information on biometric recognition accuracy.

The impact of implementing skew correction, as shown in [2], is significant. For handwriting with a skew angle of fewer than 45 degrees or exactly 45 degrees, the accuracy of the recognition system improves substantially. Specifically:

- For skew angles less than 45 degrees, the accuracy increases from 84.56% to 88.96%.
- For skew angles at 45 degrees, the accuracy improves from 23.89% to 44.11%.
- For skew angles higher than 45 degrees, the accuracy improves from 0% to 15%.

This demonstrates that correcting skew helps the convolutional neural network (CNN)-based system perform more effectively by ensuring characters are upright and aligned for better recognition during classification.

The paper [3] focuses on inclination features as an essential component of graphometric analysis for writer identification, specifically emphasizing the analysis of offline handwritten words. When inclination data is fused with other graphometric features such as proportionality index, stroke distance, and curvature, the overall classification accuracy improves dramatically, reaching 99.8–100%. This indicates that while baseline inclination alone is not highly discriminatory, it enhances system performance when combined with complementary features.

In the following subsections, we describe the normalization procedure applied to each online signature.

A. Signature Normalization

To analyze the inclination angles of online handwritten signatures, a step-by-step normalization procedure is implemented. This involves the following stages:

1) **Translation to the Center of Mass:**
   Each signature is first shifted so that its center of mass coincides with the origin. The center of mass (centroid) is given by:

$$C = (\bar{x}, \bar{y}), \quad \text{where} \quad \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{and} \quad \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i. \tag{1}$$

   Here, x_i, y_i represent the coordinates of the i-th point, and N is the total number of points.
2) **Inclination estimation:**
   Inclination was determined globally based on the second central moments of the entire signature (not on individual

strokes). Therefore, circular letters, lowercase letters, and all components of the signature contributed to the inclination estimation. We acknowledge that local variations can exist; however, since we focus on the global inclination derived from all trajectory points, the averaging process reflects the dominant orientation of the signature rather than rendering the measurement meaningless. Skew in the signature is removed by calculating the angle of skew using the second central moments:

$$\mu_{11} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}), \quad \text{(cross-covariance)} \tag{2}$$

$$\mu_{20} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2, \quad \text{(variance along the x-axis)} \tag{3}$$

$$\mu_{02} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2. \quad \text{(variance along the y-axis)} \tag{4}$$

The angle of skew θ is determined as:

$$\theta = \frac{1}{2}\arctan\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right). \tag{5}$$

The use of the MATLAB function `atan2` over `atan` in the calculation of the skew angle ensures that the full range of possible angles is accurately captured, accounting for all quadrants of the coordinate plane and thus providing a more robust and precise de-rotation of the signature.

3) **De-rotation to Correct Skew:**
A rotation matrix is applied to remove the skew. The rotation matrix is:

$$R = \begin{bmatrix}\cos(-\theta) & -\sin(-\theta)\\ \sin(-\theta) & \cos(-\theta)\end{bmatrix}. \tag{6}$$

By applying this rotation matrix, the normalized (de-skewed) coordinates (x', y') are obtained as:

$$\begin{bmatrix}x'\\ y'\end{bmatrix} = R \cdot \begin{bmatrix}x - \bar{x}\\ y - \bar{y}\end{bmatrix}. \tag{7}$$

This step removes any inclination-related skew in the signature and aligns the data along a consistent orientation. Figure 1 shows the original signature at the top, along with a straight line indicating its center of mass and inclination. At the bottom, we can observe the result of the centering and de-rotation process.

4) **Normalization:**
The normalization process standardizes the input data by removing any bias and scaling the data. The components of the signature — namely, the x, y coordinates, and p pressure values — are normalized to have a mean of zero and a standard deviation of one. This is done for each of the three variables:

$$x_{\text{norm}} = \frac{x - \mu_x}{\sigma_x}, \quad y_{\text{norm}} = \frac{y - \mu_y}{\sigma_y}, \quad p_{\text{norm}} = \frac{p - \mu_p}{\sigma_p} \tag{8}$$

where:

- μ_x, μ_y, μ_p are the mean values of the coordinates and pressure,
- $\sigma_x, \sigma_y, \sigma_p$ are their standard deviations.

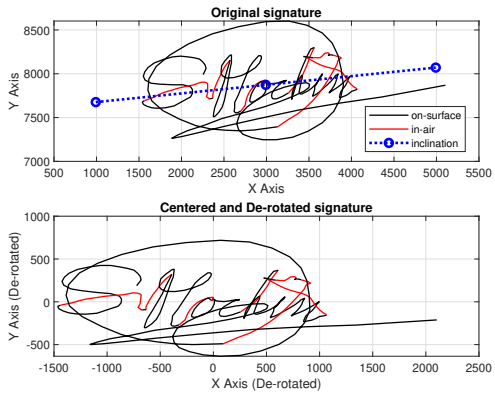


Fig. 1. Original signature with center of mass ('o') and inclination line (top), and the result after centering and de-rotation (bottom)

5) **Feature extraction:**
After normalization, the first and second derivatives of the signature are computed to extract dynamic features. The feature extraction process uses the following steps:
First Derivative (Velocity): The first derivative of the signature is calculated using a window length of 11 points [4]. It represents the velocity of the signature in the x, y, and pressure directions:

$$\dot{x} = \frac{dx}{dt}, \quad \dot{y} = \frac{dy}{dt}, \quad \dot{p} = \frac{dp}{dt} \tag{9}$$

Second Derivative (Acceleration): Similarly, the second derivative is computed to represent the acceleration:

$$\ddot{x} = \frac{d^2x}{dt^2}, \quad \ddot{y} = \frac{d^2y}{dt^2}, \quad \ddot{p} = \frac{d^2p}{dt^2} \tag{10}$$

For the code, the derivatives are computed using the 'diff()' function, which calculates the difference between adjacent points in the signature.
Normalized Derivatives: The first and second derivatives are also normalized by their respective mean and standard deviations:

$$\dot{x}_n = \frac{dx_n - \mu_{dx}}{\sigma_{dx}}, \quad \dot{y}_n = \frac{dy_n - \mu_{dy}}{\sigma_{dy}}, \quad \dot{p}_n = \frac{dp_n - \mu_{dp}}{\sigma_{dp}} \tag{11}$$

$$\ddot{x}_n = \frac{d^2x_n - \mu_{ddx}}{\sigma_{ddx}}, \quad \ddot{y}_n = \frac{d^2y_n - \mu_{ddy}}{\sigma_{ddy}}, \quad \ddot{p}_n = \frac{d^2p_n - \mu_{ddp}}{\sigma_{ddp}} \tag{12}$$

Final Feature Vector

The final feature vector $\mathbf{v}_n$ for each signature consists of:

$$\mathbf{v}_n = [x_n, y_n, p_n, \dot{x}_n, \dot{y}_n, \dot{p}_n, \ddot{x}_n, \ddot{y}_n, \ddot{p}_n] \tag{13}$$

This vector is calculated for each signature sample.

B. Experimental Setup

In our experiments we have used the MCYT database [5]

- **Signers:** N=330 individuals were analyzed.
- **Genuine Signatures:** 25 per signer.
- **Skilled Forgeries:** 25 per signer, produced by 5 individuals, ensuring diversity and avoiding bias from a single forger. It is a forgery where the impostor had access to a visual example of the genuine signature and had time to practice before attempting to replicate it.

Statistical tests were performed to assess the data distribution and differences between genuine signatures and forgeries.

II. Results

A. Inclinations

The variance of the inclinations for each signer is calculated using the formula for variance:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(\theta_i - \mu)^2 \tag{14}$$

where:

- θ_i is the inclination of the i-th signature,
- μ is the mean inclination for the signature,
- $n = 25$ is the total number of signatures per signer.

After computing the variance for each signature, the mean variance across all signers is calculated for genuine and skilled signatures (15):

$$\mu_{\text{var, type}} = \frac{1}{N}\sum_{i=1}^{N}\sigma^2_{\text{type},i}, \tag{15}$$

Similarly, the mean inclination for the entire set of signatures is defined as:

$$\mu_{\text{type}} = \frac{1}{M} \sum_{i=1}^{M} \theta_i, \tag{16}$$

where:

- type $\in \{\text{genuine, skilled}\}$ distinguishes between genuine and skilled forgeries,
- $N = 330$ is the total number of signers,
- $M = N \times 25 = 8250$ is the total number of inclination values.

Table I summarizes the results:

TABLE I
INCLINATION RESULTS FOR GENUINE SIGNATURES AND SKILLED FORGERIES.

Metric	Genuine	Skilled
μ_{var}	48.0473	80.5454
μ_{global}	9.5524°	7.6180°

Based on a statistical analysis of the inclination results for genuine signatures and skilled forgeries, we summarize the following findings:

- **Statistical Tests:** The sample distributions for at least one group are not normal. The Mann-Whitney U-test was performed, indicating a significant difference in the inclination distributions between genuine signatures and skilled forgeries. The global mean inclination (μ_{global}) for genuine signatures (9.5524) is slightly higher than that for skilled forgeries (7.6180).
- **Variance Analysis:** Levene's test was used to assess whether the variances of the two groups are equal. The test reveals a significant difference in variances between the two groups. Consequently, the mean variance (μ_{var}) for skilled forgeries (80.5454) is notably higher than that for genuine signatures (48.0473).

The histograms in Figure 2 illustrate the inclination distributions for genuine signatures and skilled forgeries.

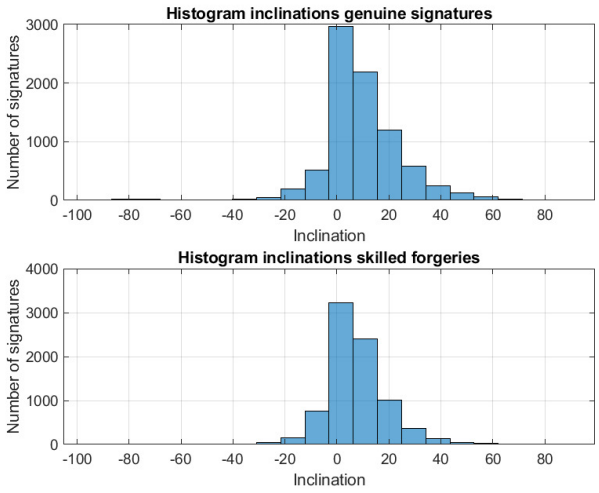


Fig. 2. Histograms of inclination angles (degrees) for genuine signatures (top) and skilled forgeries (bottom).

B. Recognition experiments

We performed identification and verification experiments both with and without inclination removal. The algorithm used to compare pairs of signatures was DTW [6] and we evaluated several different scenarios:

- 1 training signature: The recognition decision is based on a single signature template model.
- 5 training signatures: The recognition decision is based on the minimum DTW distance between the test sample and five different training signatures (data fusion at score level [7].
- Just Normalization: The signature's center of mass is shifted to (0,0) spatial coordinate
- Normalization + De-rotation: In addition to the previous normalization, the inclination of the signature is removed (the signature is de-rotated).

The experimental results are shown in table II.

TABLE II
BIOMETRIC RECOGNITION PERFORMANCE FOR IDENTIFICATION AND FORGERY SCENARIOS (RANDOM AND SKILLED).

Training signatures	Scenario	Ident. (%)	Random DCF (%)	Skilled DCF (%)
1	Normalization + De-rotation	97.2727	1.6101	5.40
	Just Normalization	98.2424	1.1506	4.9152
5	Normalization + De-rotation	99.2121	0.66971	3.3333
	Just Normalization	99.3333	0.59805	3.0242

III. CONCLUSION

The experimental results demonstrate that:

- Skilled forgeries exhibit higher variance in inclination angles compared to genuine signatures.
- The mean inclination is slightly lower for skilled forgeries, but the variability is the distinguishing factor.
- Statistical tests confirm significant differences between genuine signatures and forgeries in both inclination means and variances.
- De-rotation of signatures does not significantly affect biometric recognition accuracy (both identification and verification).

The higher variability in inclination angles observed in skilled forgeries suggests that forgers struggle to replicate this subtle yet distinctive feature. Integrating inclination variability into recognition algorithms could enhance the detection of skilled forgeries. However, this study is limited by its reliance on the MCYT database, which may not fully represent diverse populations or writing styles. Future research should validate these findings across larger and more varied datasets, exploring the inclusion of inclination features in real-time biometric systems.

The study presented here is not only significant for signature verification but also lays important groundwork for future developments in the field of watermarking of handwritten signals [8], [9]. One of the common attack strategies against watermarking systems is the application of geometric transformations, particularly rotation, which can disrupt the embedded watermark and impair subsequent recognition tasks. However, the methodology proposed in this work—particularly the characterization and normalization of the inclination axis—offers a promising solution to mitigate such attacks.

By implementing a robust de-rotation normalization step prior to recognition or authentication, future watermarking schemes can ensure that the key structural features of handwritten signals are preserved, even under adversarial conditions. Our findings suggest that inclination analysis provides a stable reference frame for this normalization, making watermark extraction and handwriting recognition resilient against rotation-based attacks. Thus, the results of this study not only contribute to signature verification but also open a path toward the design of more secure and robust watermarking algorithms for online handwriting signals.

ACKNOWLEDGEMENT

This work has been supported by Grant PID2023-146644OB-I00 funded by MICIU/AEI/10.13039/501100011033 and FEDER EU.

REFERENCES

[1] M. Faundez-Zanuy, "Biometric security technology," *IEEE Aerospace and Electronic Systems Magazine*, vol. 21, no. 6, pp. 15–26, 2006.
[2] S. Uyun, S. Rahardyan, and M. Anshari, "Skew correction and image cleaning handwriting recognition using a convolutional neural network," *International Journal on Informatics Visualization*, vol. 7, no. 3, pp. 681–687, 2023.
[3] J. L. Vásquez, A. G. Ravelo-García, J. B. Alonso, M. K. Dutta, and C. M. Travieso, "Writer identification approach by holistic graphometric features using off-line handwritten words," *Neural Computing and Applications*, vol. 32, no. 12, pp. 15 733–15 746, 2020.
[4] M. Faundez-Zanuy and M. Diaz, "On the use of first and second derivative approximations for biometric online signature recognition," in *Advances in Computational Intelligence*, I. Rojas, G. Joya, and A. Catala, Eds. Cham: Springer Nature Switzerland, 2023, pp. 461–472.
[5] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, D. Escudero, and Q.-I. Moro, "Mcyt baseline corpus: a bimodal biometric database," *IEE Proceedings - Vision, Image and Signal Processing*, vol. 150, no. 6, pp. 395–401, 2003.
[6] M. Faundez-Zanuy, "On-line signature recognition based on vq-dtw," *Pattern Recognition*, vol. 40, no. 3, pp. 981–992, 2007.
[7] ——, "Data fusion in biometrics," *IEEE Aerospace and Electronic Systems Magazine*, vol. 20, no. 1, pp. 34–38, 2005.
[8] ——, "Comprehensive analysis of least significant bit and difference expansion watermarking algorithms for online signature signals," *Expert Systems with Applications*, vol. 267, p. 126214, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417424030811
[9] ——, "Online signature watermarking in the transform domain," *Cognitive Computation*, vol. 17, no. 2, p. 79, 2025. [Online]. Available: https://doi.org/10.1007/s12559-025-10436-y

Quasi-Synthetic Riemannian Data Generation for Writer-Independent Offline Signature Verification

Elias N. Zois[(a)], Moises Diaz[(b)], Salem Said[(c)], Miguel A. Ferrer[(b)]

[a]*University of West Attica, Aigaleo, Greece.* [b]*Instituto Universitario para el Desarrollo Tecnologico y la Innovacion en Comunicaciones. Universidad de Las Palmas de Gran Canaria, Campus de Tafira, Spain.* [c]*Université Grenoble-Alpes, Laboratoire Jean Kuntzmann*

[(a)]ezois@uniwa.gr, [(b)]{moises.diaz, miguelangel.ferrer}@ulpgc.es, [(c)]salem.said@univ-grenoble-alpes.fr

***Abstract*—** Offline handwritten signature verification remains a challenging task, particularly in writer-independent settings where models must generalize across unseen individuals. Recent developments have highlighted the advantage of geometrically inspired representations, such as covariance descriptors on Riemannian manifolds. However, past or present, handcrafted or data-driven methods usually depend on real-world signature datasets for classifier training. We introduce a quasi-synthetic data generation framework leveraging the Riemannian geometry of Symmetric Positive Definite matrices (SPD). A small set of genuine samples in the SPD space is the seed to a Riemannian Gaussian Mixture which identifies Riemannian centers as synthetic writers and variances as their properties. Riemannian Gaussian sampling on each center generates positive as well as negative synthetic SPD populations. A metric learning framework utilizes pairs of similar and dissimilar SPD points, subsequently testing it over on real-world datasets. Experiments conducted on two popular signature datasets, encompassing Western and Asian writing styles, demonstrate the efficacy of the proposed approach under both intra- and cross- dataset evaluation protocols. The results indicate that our quasi-synthetic approach achieves low error rates, highlighting the potential of generating synthetic data in Riemannian spaces for writer-independent signature verification systems.

Keywords*— *Metric learning, Riemannian Gaussian Distributions, Symmetric positive definite manifold, Writer independent offline signature verification

I. Introduction

Handwriting, an outcome of the human movement, remains a fundamental human skill [1], playing a significant role in various forensic and security applications, despite the rise of digital technologies. Among handwriting-based authentication methods, handwritten signature verification presents a complex challenge [2], particularly in writer independent (WI-SV) operation where methods must generalize across unseen writers [3]. Typically, WI-SV models are learned (i.e. Trained and validated) on a development set [4] and then exploited (or tested) on blind datasets, following intra-lingual or cross-lingual evaluation protocols. On intra-lingual protocol, half of the writers in a dataset should be utilized for learning and the other half for testing and then, these subsets exchange roles. On cross-lingual protocol, the WI classifier is learned over an entire dataset and then tested on different datasets [5], [6]. Traditional WI-SV approaches involve a combination of "handcrafted" and/or "data-driven" techniques, with examples ranging from the dichotomy transform [7] to contemporary deep learning topologies [8].

Most of these cases require extensive real-world signature datasets for training. Also, methodologies employing synthetic data augmentation techniques have been proposed primarily with the GPDSsynthetic dataset [9] for model learning [10] or with the Signet-S [11], a popular deep-learning feature extraction module. Another line of research addresses augmentation techniques in the feature (or descriptor) domain [3], [12]. This solution is motivated by the fact that more important details can be found by analyzing extracted signature features rather than examining the signature image itself. However one may identify the absence of a probabilistic model that is both rigorously defined and tractable, thereby enabling the representation of the statistical variability of data. Recent studies have explored the potential of leveraging matrix manifolds for the purpose of mapping a signature image through its covariance

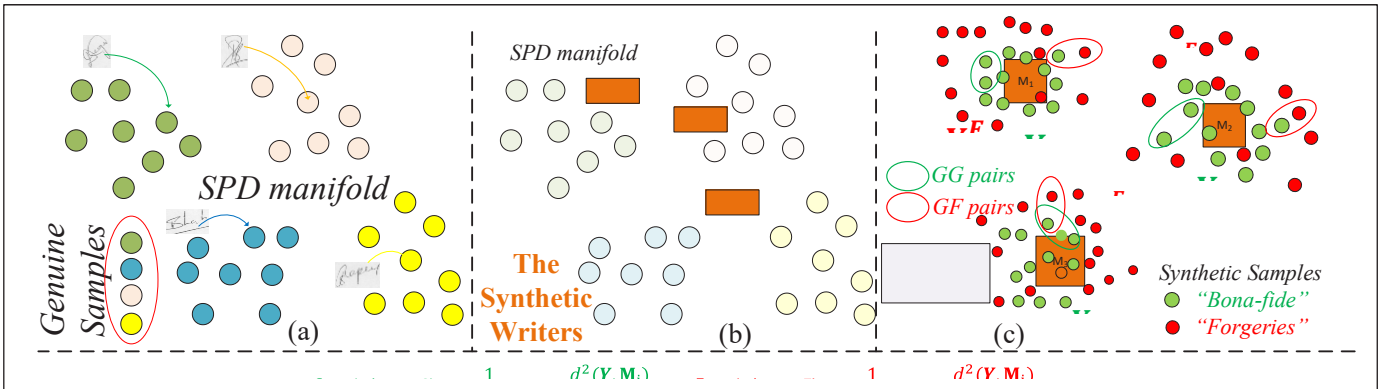


Fig. 1. Toy example of the proposed quasi-synthetic data generation. a) Out of $\mathfrak{M} = 4$ writers and $\mathfrak{D} = 8$ samples per writer we have a set of $\mathfrak{M} \cdot \mathfrak{D} = 32$ handwritten signatures and their $\boldsymbol{X} = \{\boldsymbol{X}_{t=1:32}\}$ SPD representations. b) A R-GMM on $\boldsymbol{X}$ returns three SPD centers $\boldsymbol{M}_i$ along with their corresponding variances σ_i, $i = 1{:}3$. c) For each $\boldsymbol{M}_i$ we draw a number of $L = 12$ genuine $\boldsymbol{Y}_i^G$ and forgery $\boldsymbol{Y}_i^F$ synthetic SPD points out of two Riemannian Gaussian Distributions $\boldsymbol{Y}_i^{G,F} \sim G(\boldsymbol{M}_i, \sigma_i^{G,F})$ in which $\sigma_i^G = \sigma_i$, $\sigma_i^F = a \cdot \sigma_i$, $a > 1$. The metric learning based classifier will be developed using similar $\boldsymbol{Y}_i^G \boldsymbol{Y}_i^G$ (GG) pairs and dissimilar $\boldsymbol{Y}_i^G \boldsymbol{Y}_i^F$ (GF) pairs.

descriptor to the symmetric positive definite (SPD) manifold [5], [6], [13] defined as the set of all $\boldsymbol{Y} \in \mathbb{R}^{d\times d}$ real matrices: $\boldsymbol{Y} \in \boldsymbol{P}_d, \boldsymbol{Y^T - Y = 0}^{d\times d}, \boldsymbol{v^T Y v} > 0, \forall \boldsymbol{v} \in \mathbb{R}^d - \boldsymbol{0}^d$. The "image to SPD matrix" encoding is achieved by its associated covariance matrix of d image filters as a region descriptor. To this end, reference [14] introduced the framework for Riemannian Gaussian Distributions (RGD), while reference [15] used it for writer dependent (WD) signature verification.

In this study, we explore a Riemannian Gaussian Mixture Model (R-GMM) [14] as a novel approach to identify a number of K-synthetic writers accompanied by their statistical parameters, i.e. centers $\{\boldsymbol{M}_{i=1:K} \in \boldsymbol{P}_d\}$ and variances $\{\sigma_i\}$. The seed to R-GMM is a small population of genuine samples. The take-home message of the proposed article is about: a) modeling handwritten signature properties to covariance matrices – SPD points, b) create, under a soft clustering procedure, synthetic writers and c) duplicate new samples for improvement of the learning procedure. The synthetics $\boldsymbol{M}_i$ are used to randomly draw synthetic SPD matrices of two kinds: In the first one, for each center $\boldsymbol{M}_i$ we draw SPD samples, which mimic genuine samples, with $\sigma_i^G = \sigma_i$, while in the second one, we draw SPD samples with $\sigma_i^F = a \cdot \sigma_i, a > 1$ (where a is a scaling factor), in order to generate synthetic forgeries of varying qualities. Finally, an SPD-related metric learning algorithm [6] is employed for offline signature verification on two popular datasets.

II. The Proposed Approach

A. The big picture

Fig. 1 presents a simple toy example in which we visually demonstrate the proposed method. Given a development set of $\mathfrak{M}$-users with $\mathfrak{D}$-development samples per user, expressed by their corresponding SPD manifold set $\boldsymbol{X} = \{\boldsymbol{X}_{t=1:\mathfrak{M}\cdot\mathfrak{D}}\} \in \boldsymbol{P}_d$ the R-GMM algorithm [14] identifies K synthetic writers and returns the model $\theta = \{\omega_i, \boldsymbol{M}_i, \sigma_i\}, i = 1{:}K$ with $\omega_i, \boldsymbol{M}_i, \sigma_i$ to be the corresponding non-zero positive weights, means and variances. Each $\boldsymbol{M}_i$ center (a synthetic writer in the SPD manifold) assimilates visual attributes from the set $\{\boldsymbol{X}_t\}$. Following the estimation of the $\boldsymbol{M}_i$ synthetic writers we do not need anymore the human-genuine samples. Synthetic genuine $\boldsymbol{Y}_i^G$ and forgeries $\boldsymbol{Y}_i^F$ SPD points are artificially draw from $\boldsymbol{Y}_i^G \sim G(\boldsymbol{M}_i, \sigma_i^G)$ and $\boldsymbol{Y}_i^F \sim G(\boldsymbol{M}_i, \sigma_i^F)$ RGDs. In the development stage of an WI-SV verifier, similar $(\boldsymbol{Y}_i^G \boldsymbol{Y}_i^G)$ as well as dissimilar $(\boldsymbol{Y}_i^G \boldsymbol{Y}_i^F)$ SPD pairs are employed in batches.

B. Mixtures of RGD: Parameter estimation and generation of synthetic data

An RGD $\boldsymbol{Y} \sim G(\boldsymbol{M}, \sigma)$ on the SPD manifold $\boldsymbol{P}_d$ resembles the typical $G(\mu, \sigma)$ in the Euclidean space $\mathbb{R}^d$. It is defined by a probability density function (p.d.f) $\boldsymbol{Y} \sim G(\boldsymbol{M}, \sigma)$ where $\mathbf{M} \in \boldsymbol{P}_d$ is the Riemannian mean, $\sigma \in \mathbb{R}_+$ express the dispersion, $d^2(\boldsymbol{Y}, \mathbf{M})$ is Rao's distance and $\zeta(\sigma)$ is a normalizing factor [14].

$$\boldsymbol{Y} \sim G(\boldsymbol{M}, \sigma) = \frac{1}{\zeta(\sigma)} \exp\left[-\frac{d^2(\boldsymbol{Y}, \mathbf{M})}{2\sigma^2}\right] \quad (1)$$

Mixtures of RGDs (R-GMM) is also a probabilistic distribution on $\boldsymbol{P}_d$ with p.d.f as of (2) where K represents the total number of mixture components (in our case the synthetic writers) [14]:

$$p(\boldsymbol{Y}) = \sum_{i=1}^{K} \omega_i \times G(\boldsymbol{M}_i, \sigma_i) \quad (2)$$

Prior to the utilization of the R-GMM algorithm, we map all signature images to the corresponding covariance matrices $\boldsymbol{X}_t \in \boldsymbol{P}_d$; in our case $d = 10$, i.e. the manifold dimensionality is set to ten. The preprocessing steps as well as the creation of the covariance matrices $\{\boldsymbol{X}_t\}$ are addressed analytically in the relative literature [5], [6]. The computation of the parameters of the R-GMM model $\theta = \{\omega_i, \boldsymbol{M}_i, \sigma_i\}$ is achieved through a novel expectation maximization (EM) algorithm in the SPD manifold [14]. Any $\boldsymbol{M}_i$ center along with its corresponding variance σ_i is now employed in order to generate a population of synthetic SPD points $\boldsymbol{Y}_i^G, \boldsymbol{Y}_i^F$ according to (1) and Fig. 1. In our case for each center $\boldsymbol{M}_i$, a total of $L = 15$ $\boldsymbol{Y}_i^G \in \boldsymbol{P}_{10}$ and $\boldsymbol{Y}_i^F \in \boldsymbol{P}_{10}$ matrices is randomly draw. To control the dispersion of synthetic $\boldsymbol{Y}_i^G, \boldsymbol{Y}_i^F$ points from the $\boldsymbol{M}_i$ we utilize two design dispersion parameters namely: a) $\sigma_i^G = \sigma_i$ for the synthetic genuine and b) $\sigma_i^F = a \cdot \sigma_i$ for the synthetic forgeries, in which the scaling factor a controls the dispersion of $\boldsymbol{Y}_i^F \sim \mathrm{G}(\boldsymbol{M}_i, \sigma_i^F)$.

In this work the scaling factor a was set to a number of two ad-hoc values: $a = 1.1$ or $a = 2$ in order to intuitively identify two diverse forgery qualities: in the first the generated $\boldsymbol{Y}_i^F$ samples are considered to be proximal to the $\boldsymbol{Y}_i^G$, while in the second case they are apart. Therefore, during the development stage of the classifier and for $a = 1.1$ dissimilar $\boldsymbol{Y}^G \boldsymbol{Y}^F$ pairs will tend to exhibit $d^2(\boldsymbol{Y}_i^G, \boldsymbol{Y}_i^F)$ distances that significantly overlap with the $d^2(\boldsymbol{Y}_i^G, \boldsymbol{Y}_i^G)$ distances, contrary to the $a = 2$ case. We consider the $a = 1.1$ case to correspond to the scenario in which the ω^- class is represented by pairs of real-world genuine and skilled forgeries (GSF) while the $a = 2$ case corresponds to the scenario in which the ω^- class is represented by pairs of genuine and other people's genuine (or random forgeries, GRF). For simplicity we refer to the $a = 1.1$ and the $a = 2$ cases as "Hard" or "Soft" training.

C. The metric learning classifier

We address the WI-SV classifier as a binary SPD metric learning problem denoted as $\Delta = \{\{\boldsymbol{W}_c\}, A, \boldsymbol{\Sigma}\}$ with $\{\boldsymbol{W}_c\}, A, \boldsymbol{\Sigma}$ its design parameters [6]. In details, for any given pair of SPD matrices $\boldsymbol{Y}_i^G \boldsymbol{Y}_i^G$ or $\boldsymbol{Y}_i^G \boldsymbol{Y}_i^F$ the metric learning evaluates a distance Δ as a measure of (dis)similarity between $\boldsymbol{Y}_i^G \boldsymbol{Y}_i^G$ or $\boldsymbol{Y}_i^G \boldsymbol{Y}_i^F$. Ideally, Δ should be low for similar $\boldsymbol{Y}_i^G \boldsymbol{Y}_i^G$ and high for dissimilar $\boldsymbol{Y}_i^G \boldsymbol{Y}_i^F$ pairs. Some details regarding the learnable model parameters follow:

- $\boldsymbol{W}_c \in \mathbb{R}^{d\times q}$ with $c = 1{:}\,m$ is a set of orthogonal point-to-set transformation matrices which projects any high dimensional SPD point $\boldsymbol{Y}_i, \in \boldsymbol{P}_d$ to a set of m, low dimensional SPD matrices $\{\boldsymbol{Y}_i^c\} \in \boldsymbol{P}_q$, $d = m \cdot q$ by the congruence $\boldsymbol{Y}_i^c = \boldsymbol{W}_c^T \boldsymbol{Y}_i \boldsymbol{W}_c$. Thus pairs of high dimensional SPD points $\boldsymbol{Y}_i^G, \boldsymbol{Y}_i^{G,F} \in \boldsymbol{P}_d$ are converted to a total of m sub-SPD pairs $\boldsymbol{Y}_i^{G,\{c\}}, \boldsymbol{Y}_i^{G,F,\{c\}} \in \boldsymbol{P}_q$.
- $A \in \mathbb{R}^{m\times 2}$ is a matrix comprised by a set of $(\alpha_c, \beta_c)_{c=1}^m$ values. They are the learnable parameters of the alpha-beta divergence on each low dimensional manifold point $\boldsymbol{Y}_i^c \in \boldsymbol{P}_q$. For a $\boldsymbol{Y}_i^G, \boldsymbol{Y}_i^{G,F}$ pair on $\boldsymbol{P}_d$, an arbitrary $\boldsymbol{Y}_i^{G,c}, \boldsymbol{Y}_i^{G,F,c}$ pair on $\boldsymbol{P}_q$ and two scalars (α_c, β_c) the $d^{(\alpha_c,\beta_c)}\left(\boldsymbol{Y}_i^{G,c} \middle\| \boldsymbol{Y}_i^{G,F,c}\right)$ is a positive real number according to:

$$d_c^{(\alpha_c,\beta_c)}\left(\boldsymbol{Y}_i^{G,c} \middle\| \boldsymbol{Y}_i^{G,F,c}\right) = \frac{1}{\alpha_c \cdot \beta_c} \log\left(\det\left(\frac{\alpha_c + \beta_c\left(\boldsymbol{Y}_i^{G,c}(\boldsymbol{Y}_i^{G,F,c})^{-1}\right)^{-\alpha_c}}{\alpha_c + \beta_c}\right)\right) \qquad (3)$$

- $\boldsymbol{\Sigma} \in \boldsymbol{P}_m$ is an SPD matrix entailed to map the embedding vector $\boldsymbol{v} = \left\{d_c^{(\alpha_c,\beta_c)}\right\} \in \mathbb{R}^m$ to the final $\Delta \in \mathbb{R}$ distance: $\Delta = \boldsymbol{v}^T \boldsymbol{\Sigma} \boldsymbol{v}$. For simplicity reasons, this work employ the $q = 10$, $m = 1$ values. Thus, $c = 1$, $A \in \mathbb{R}^2$ and therefore $\boldsymbol{\Sigma}$ is obsolete.

III. Experimental Setup And Results

We demonstrate the effectiveness of the proposed method for the CEDAR (55 writers, 24 genuine and skilled forgeries for each writer) and the Bengali subset of the BHSig260 database (100 writers, 24 genuine and 30 skilled forgeries for each writer). We design an intra-lingual F_{intra} and a cross-lingual F_{cross} experiment in which the development and the testing sets were created as follows [5], [6] [13]: Intra-lingual (F_{intra}): The classifier is developed on 50% of the writers and tested on the other 50%. Specifically, CEDAR: 27 writers for development, 28 for testing. Roles were reversed for a second evaluation. Bengali: 50 writers for development, 50 for testing. Roles were reversed similarly. Experiments were repeated 5 times (i.e. a 5×2 fold). Cross-lingual (F_{cross}): The classifier is developed on the 100% of the writers of one dataset and tested/transferred on the 100% of the other.

For each development writer we select a number of $\mathfrak{D} = 7$ genuine samples. Then, the R-GMM of (2) is applied for $K = 3,7,11,15$ centers. Next, for each center we generate $L = 15$ synthetic samples according to the procedure described in section II and Fig. 1 for $\boldsymbol{Y}_i^G$ and $\boldsymbol{Y}_i^F$, $i = 1{:}\,K$ and for both cases of the scaling factor $a = 1.1$ (Hard) or $a = 2$ (Soft). Then, pairs of $\boldsymbol{Y}_i^G \boldsymbol{Y}_i^G$, $\boldsymbol{Y}_i^G \boldsymbol{Y}_i^F$ are selected in order to learn and validate the $\Delta = \{\{\boldsymbol{W}_c\}, A, \boldsymbol{\Sigma}\}$ distance. The testing stage of [5], [6] [13] (repeated ten times) applies:

- For each writer of the testing set ten genuine samples are reserved as the reference image set $\{\mathcal{R}\}$ and SPD points $\{\boldsymbol{\mathcal{R}}\}$. The remaining genuine (ω^+) and skilled forgeries (ω^-) form the questioned SPD set $\boldsymbol{Q}^{\pm}$ of size $|\boldsymbol{Q}^{\pm}|$. For each $\mathcal{R}$, Q images an array of fourteen covariance matrices $\boldsymbol{\mathcal{R}}_g$, $\boldsymbol{Q}_g$, $g = 1{:}\,14$ is created as a result of a spatial equimass pyramid image partition. In detail, $\boldsymbol{\mathcal{R}}_1$ points to the SPD matrix of the whole image, $\boldsymbol{\mathcal{R}}_2, \ldots, \boldsymbol{\mathcal{R}}_5$ points to the four SPD matrices of a 2×2 equimass image segmentation and $\boldsymbol{\mathcal{R}}_6, \ldots, \boldsymbol{\mathcal{R}}_{14}$ points to the nine SPD matrices of a 3× 3 equimass image segmentation. Thus, for a pair of SPD points $(\boldsymbol{\mathcal{R}}, \boldsymbol{Q}^{\pm})$ fourteen SPD pairs $\left(\boldsymbol{\mathcal{R}}_g, \boldsymbol{Q}_g^{\pm}\right)$, $\mathrm{g} = 1{:}\,14$ are used as the input to the Δ distance; consequently this forms a corresponding vector of scores $\Delta_{\pm} \in \mathbb{R}^{14}$.
- The resulted $\Delta_{\pm}$ vector derived by a $(\boldsymbol{\mathcal{R}}, \boldsymbol{Q}^{\pm})$ pair is further sorted (i.e. $s\Delta_{\pm}$) and next re-arranged as $\Delta_{u\pm} \in \mathbb{R}^{14}$: $\Delta_{u\pm}(g) = mean(s\Delta_{\pm}(1{:}\,g))$. Consequently, for all questioned and reference pairs $(\{\boldsymbol{\mathcal{R}}\}, \boldsymbol{Q}^{\pm})$, a tensor $T_{Q^{\pm}} \in \mathbb{R}^{|Q^{\pm}|\times 10\times 14}$ accumulates the total scores. By taking the minimum over all ten references, a final table of scores $T_{sc_{Q^{\pm}}} \in \mathbb{R}^{|Q^{\pm}|\times 14}$ is formed. We explore each g-column of $T_{sc_{Q^{\pm}}}$ by a Receiver Operating Characteristic curve in order to evaluate the equal error rate (EER_{SF} %) per each testing writer, and per the parameter g. For each writer this is repeated ten times; averages are reported.

Fig. 2 display the average EER_{SF} (%), plotted as a function of the parameter g. The results indicate that the proposed quasi-synthetic approach achieves low verification error rates, demonstrating its effectiveness in reducing reliance on real-world datasets. Beside this, two key findings emerge from the comparison of the EER_{SF} curves. First, the results show that the "hard" training protocol consistently outperforms the "soft" training protocol across both datasets and for both intra-dataset (F_{intra}) and cross-dataset (F_{cross}) evaluation protocols. Within the "hard" training paradigm, the (F_{cross}) protocol demonstrates slightly better performance than the (F_{intra}) one. Second, regarding the number of clusters, Fig. 2 suggests that K=7 is a favorable choice. However, the overall system performance appears largely independent of K, as the error curves remain relatively stable across different values of it. A slight variation is observed when training on 100% of the CEDAR and testing on 100% of the Bengali, where lower values of yield slightly better results. In summary, our approach achieves competitive results, with EER values consistently below 0.4% across all studied cases. These promising outcomes encourage further exploration of additional configurations to validate the stability of our findings.

IV. Conclusions

This work introduces a Riemannian quasi-synthetic data generation framework for writer-independent offline signature verification. Through Riemannian Gaussian Mixture models, we synthesize SPD representations that reduce dependency on extensive

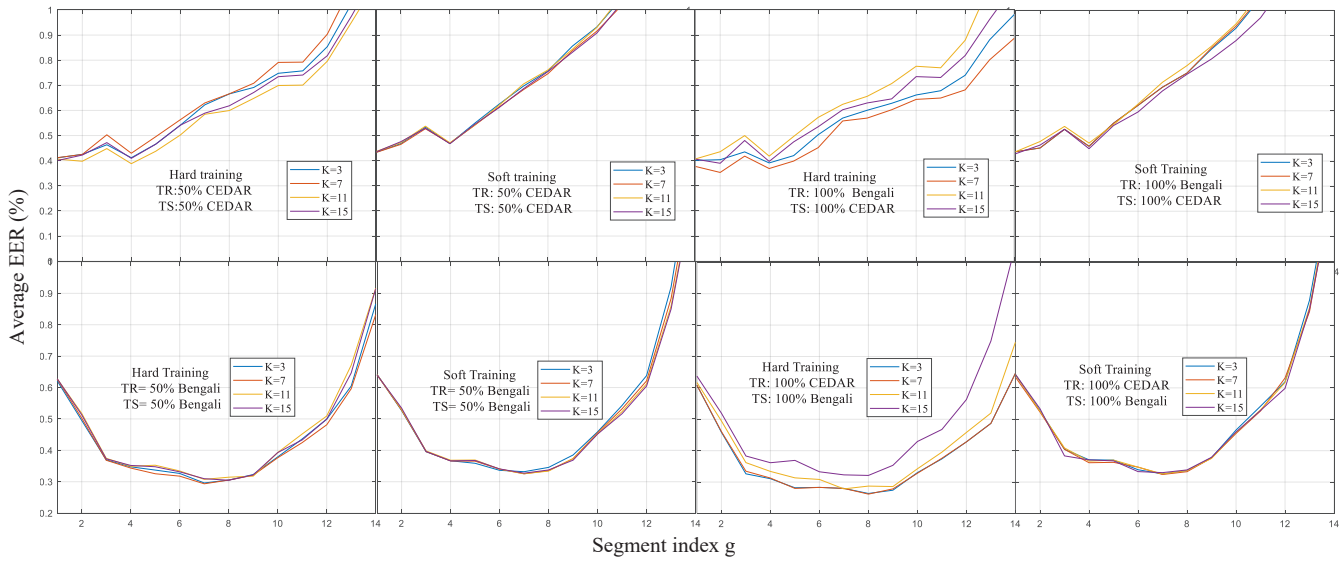


Fig. 2. Average EER_{SF} (%) as a function of the index g for: the F_{intra}, F_{cross}, "Hard" & "Soft", protocols $L = 15$, and $K = 3,7,11,15$ centers. TR, TS denotes Training/Testing datasets.

real-world datasets while maintain competitive verification performances. Our findings highlight the potential of leveraging synthetic data generation techniques in advancing robust and scalable signature verification systems.

ACKNOWLEDGMENT

This research was partly supported by the PID2023-146620OB-I00, funded by MICIU/AEI 0.13039/501100011033 and the European Union's FEDER program, partly by the CajaCanaria and la Caixa (2023DIG05) and partly by the University of West Attica, Greece.

REFERENCES

[1] M. Diaz, A. Mendoza-García, M. A. Ferrer and R. Sabourin, "A survey of handwriting synthesis from 2019 to 2024: A comprehensive review," *Pattern Recognition,* p. 111357, 2025, https://doi.org/10.1016/j.patcog.2025.111357.

[2] M. Diaz, M. A. Ferrer, D. Impedovo, M. I. Malik, G. Pirlo and R. Plamondon., "A Perspective Analysis of Handwritten Signature Technology," *ACM Computing Survey 51, 6, Article 117 (January 2018),* pp. 1-39, 2018, doi.org/10.1145/3274658.

[3] N. Arab, H. Nemmour and Y. Chibani, "A new synthetic feature generation scheme based on artificial immune systems for robust offline signature verification," *Expert Systems with Applications,* vol. 213, p. 119306, 2023, doi.org/10.1016/j.eswa.2022.119306.

[4] D. Rivard, E. Granger and R. Sabourin, "Multi-feature extraction and selection in writer-independent off-line signature verification," *International Journal on Document Analysis and Recognition,* vol. 16, no. 1, pp. 83-103, 2013, doi: 10.1007/s10032-011-0180-6.

[5] A. Giazitzis and E. N. Zois, "Metric meta-learning and intrinsic Riemannian embedding for writer independent offline signature verification," *Expert Systems with Applications,* vol. 261, p. 125470, 2025, doi.org/10.1016/j.eswa.2024.125470.

[6] E. N. Zois, D. Tsourounis and D. Kalivas, "Similarity Distance Learning on SPD Manifold for Writer Independent Offline Signature Verification," *IEEE Transactions on Information Forensics and Security,* vol. 19, pp. 1342-1356, 2024, doi: 10.1109/TIFS.2023.3333681.

[7] S.-H. Cha and S. Srihari, "Writer Identification: Statistical Analysis and Dichotomizer," in *Advances in Pattern Recognition*, vol. 1876, F. Ferri, J. Iñesta, A. Amin, and P. Pudil Eds., (Lecture Notes in Computer Science: Springer Berlin Heidelberg, 2000, ch. 13, pp. 123-132.

[8] L. Zheng, D. Wu, S. Xu and Y. Zheng, "HTCSigNet: A Hybrid Transformer and Convolution Signature Network for offline signature verification," *Pattern Recognition,* p. 111146, 2024, doi.org/10.1016/j.patcog.2024.111146.

[9] M. A. Ferrer, M. Diaz, C. Carmona-Duarte and A. Morales, "A Behavioral Handwriting Model for Static and Dynamic Signature Synthesis," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 39, no. 6, pp. 1041-1053, 2017, doi: 10.1109/TPAMI.2016.2582167.

[10] H. Li, P. Wei, Z. Ma, C. Li and N. Zheng, "TransOSV: Offline Signature Verification with Transformers," *Pattern Recognition,* vol. 145, p. 109882, 2024, doi.org/10.1016/j.patcog.2023.109882.

[11] K. G. de Moura, R. M. O. Cruz and R. Sabourin, "Offline Handwritten Signature Verification Using a Stream-Based Approach," Cham, 2025: Springer Nature Switzerland, in (ICPR) 2024, pp. 271-286.

[12] M. M. Yapıcı, A. Tekerek and N. Topaloğlu, "Deep learning-based data augmentation method and signature verification system for offline handwritten signature," *Pattern Analysis and Applications,* vol. 24, no. 1, pp. 165-179, 2021/02/01 2021, doi: 10.1007/s10044-020-00912-6.

[13] A. Giazitzis, M. Diaz, E. N. Zois and M. A. Ferrer, "Janus-Faced Handwritten Signature Attack: A Clash Between a Handwritten Signature Duplicator and a Writer Independent, Metric Meta-learning Offline Signature Verifier," Cham, 2024: Springer Nature Switzerland, in ICDAR 2024, pp. 216-232.

[14] S. Said, L. Bombrun, Y. Berthoumieu and J. H. Manton, "Riemannian Gaussian Distributions on the Space of Symmetric Positive Definite Matrices," *IEEE Transactions on Information Theory,* vol. 63, no. 4, pp. 2153-2170, 2017, doi: 10.1109/TIT.2017.2653803.

[15] E. N. Zois, S. Said, D. Tsourounis and A. Alexandridis, "Subscripto multiplex: A Riemannian symmetric positive definite strategy for offline signature verification," *Pattern Recognition Letters,* vol. 167, pp. 67-74, 2023, doi.org/10.1016/j.patrec.2023.02.002.

Associating Velocity Profiles to Ink Traces for Forensic Handwriting Examination

Rosanna Coccaro
Automatic Control Group, DIEM
University of Salerno
rcoccaro@unisa.it

Enrico Ferrentino
Automatic Control Group, DIEM
University of Salerno
eferrentino@unisa.it

Antonio Parziale
NCLab, DIEM
University of Salerno
anparziale@unisa.it

Angelo Marcelli
NCLab, DIEM
University of Salerno
amarcelli@unisa.it

Abstract- In forensic handwriting examination, the handwriting is available on paper, and thus the kinematics of the handwriting movement is partially lost. The lost information, namely the velocity profile, is paramount for accurately characterizing the samples under investigation and their comparison. We address this issue by reformulating the problem as a time-optimal trajectory planning and introducing an algorithm that assigns velocity profiles to the offline data. We report the results of experiments to evaluate to which extent the velocity profile produced by our algorithm resembles those obtained from the online data and discuss how and to which extent the estimated kinematics can be used to identify the source of handwritten samples.

Index Terms

forensic handwriting examination, velocity profile, time-optimal trajectory planning

I. Introduction

Findings in neuroscience and human motor control have shown that handwriting is a complex human movement that involves the proper coordination of the nervous system (both central and peripheral) and the musculoskeletal one. Handwriting skills are learned through a trial-and-error process to produce the intended movement with the minimum energy consumption. Once handwriting skills are mastered, handwriting is executed by recruiting the competent musculoskeletal subsystems and timely activating them. In handwriting, the movement of interest is the *trajectory* of the tip of the pen on the writing surface, which can be described by the time-stamped sequence of points reached while drawing. Thus, this sequence represents both the *path* of the movement in space and the *time-law* of the path following during the writing. As a consequence, the individuality of handwriting translates into the individuality of trajectories: no two individuals may produce the same trajectories to execute the same intended movements, and no individual can produce the same trajectory twice.

In forensic handwriting examination, only the path is available for comparison, because handwritten samples are ink traces on a writing surface. This loss of information may cause serious consequences on the final decision because similar paths may have been executed with different velocities, as happens when a subject attempts to imitate someone else's handwriting. To pursue the goal, the forger will produce many attempts until the shape of the path satisfactorily resembles the genuine. However, as the learning is focused on the shape, the forger will use his/her motor skills, which are different from those of the author of the genuine sample. Those differences, which result from the differences in the recruitment and coordination of the musculoskeletal system used to execute the movement, translate into different forces applied to the joints. As different forces are applied, the acceleration and velocity of the pen tip differ from those adopted by the genuine author. Thus, the amount of variation may provide a robust bias for the final decision, even when the paths look similar. When this information is lost, the examiners must establish the relevance of the path similarity based on their expertise. As the truth is not available, in case two examiners adopt contrasting evaluations about the relevance of path similarity, there is no way to decide which one is valid. The loss of the time-related information also explains the results reported in the literature in automatic writer and signature verification, showing that when time-derived information is available the achieved results are better than those achieved when such information is unavailable, despite the large variety of time-related features and classifier used to reach the final decision [1], [2].

We aim to fill this gap by proposing an algorithm, based on a methodology largely adopted in robot motion planning, capable of assigning a velocity profile to a given path. We will show by experiments how we assessed the performance of our algorithm and the achieved results, and how comparing the velocity profiles produced by our algorithm on genuine and questioned samples may provide useful hints to the examiners for weighing the relevance assigned to the shape similarity of the paths in the final decision.

The remainder of this paper is organized as follows. In Section 2 we discuss the rationale behind the design of the framework for assessing its performance, while Section 3 reports the achieved results and discusses how and to which extent they may provide useful information for forensic handwriting examination. In the Conclusion, we summarize the results and outline our future work on the topic.

II. Methods and materials

As mentioned in the Introduction, the ultimate aim of our work is to enrich the representation of the handwriting by associating a synthetic velocity profile to an ink trace, so that it resembles as much as possible the lost original profile.

A. Phase plane analysis

The phase plane analysis is a typical approach in robotics used to isolate time laws and analyze them regardless of the underlying geometrical path, which can be defined either in joint space or in task space. In the latter case, which is the case of handwriting, the path is $\mathbf{x}(\lambda) \in \mathbb{R}^m$, where λ is the path parameter and, since during a handwriting pattern the pentip cannot stop or reverse its motion, it is a strictly monotonic function of time:

$$\lambda(t) : \mathbb{R} \to [0; L], \quad \dot{\lambda} = \frac{d\lambda}{dt} > 0 \quad \forall t. \tag{1}$$

The function $\lambda(t)$, or equivalently, its inverse $t(\lambda)$, represents the time law, and it has to be found as a result of the time parametrization process. The phase plane analysis deals with determining $\lambda(t)$ indirectly, through the definition of curves in the λ-$\dot{\lambda}$ plane, i.e. plotting the pseudo-velocity $\dot{\lambda}$ with respect to the path parameter λ. $\dot{\lambda}(\lambda)$ is termed *Phase Plane Trajectory* (PPT), from which the time can be computed as

$$t(\lambda) = \int_0^L \frac{1}{\dot{\lambda}(l)} dl. \tag{2}$$

Once a time law is available, it can be assigned to a path to yield a *trajectory*, i.e. $\mathbf{x}\big(\lambda(t)\big) = \mathbf{x}(t)$. Let $\mathbf{x}'(\lambda) = \frac{d\mathbf{x}(\lambda)}{d\lambda}$ and $\mathbf{x}''(\lambda) = \frac{d^2\mathbf{x}(\lambda)}{d\lambda^2}$. By using the chain rule on $\dot{\mathbf{x}}\big(\lambda(t)\big)$ and $\ddot{\mathbf{x}}\big(\lambda(t)\big)$, we obtain the equations

$$\dot{\mathbf{x}} = \mathbf{x}'\dot{\lambda}, \tag{3}$$

$$\ddot{\mathbf{x}} = \mathbf{x}'\ddot{\lambda} + \mathbf{x}''\dot{\lambda}^2, \tag{4}$$

which link the task velocity and acceleration $\dot{\mathbf{x}}(t)$ and $\ddot{\mathbf{x}}(t)$ with their parametric velocity and acceleration $\mathbf{x}'(\lambda)$ and $\mathbf{x}''(\lambda)$. To this respect, we note that the pseudo-velocity $\dot{\lambda}$ can also be regarded as the magnitude of the tangential velocity $\|\dot{\mathbf{x}}(t)\|$ when given with respect to time.

B. The STOTPAC algorithm

As planning in the phase plane amounts to defining the PPT $\dot{\lambda}(\lambda)$, we introduce a variant of the Time-Optimal Trajectory Planning (TOTP) algorithm [3], called *Scaled TOTP with Acceleration Constraints (STOTPAC)*. It incorporates the TOTP algorithm, which aims at finding $\dot{\lambda}(\lambda)$ minimizing the trajectory tracking time, subject to *acceleration constraints*. The actual values of the constraints can be derived from studies on handwriting movements or extrapolated from handwriting samples, depending on the available data. It is worth mentioning that different values of these limits affect the total time to execute the movement, but not the shape of the profiles. In our implementation, the velocity in the phase plane is *scaled* so that the execution time of the reconstructed trajectory is equal to a predefined value.

III. Performance Assessment

Our experimental work is to associate a velocity profile to the handwriting ink trace, with the ultimate aim of comparing a genuine specimen versus a questioned one. So our experiments were designed to ascertain to which extent the reconstructed profile resembles the original ones, and if the intra-variability observed on the original profiles is preserved in the reconstructed ones.

A. Data set

We collected handwriting samples produced by 8 subjects, 7 male and 1 female, aged between 23 and 33 years (mean 26.9, std ± 3.9) who were requested to draw 10 times each of the 3 handwriting patterns represented in Figure 1 using an ink pen and a paper sheet fixed on top of a WACOM Intuos 2 tablet with a sampling rate of 100 Hz. This setup allowed us to simultaneously record the trajectory provided by the tablet and the corresponding ink trace in digital format by scanning the paper sheet. Motion data of the pen on the 2D tablet plane was recorded using MovAlyzeR v6.1 [4] to obtain the velocity profile of the original trajectory, hereinafter denoted as *Human*, and compute maximum and minimum acceleration, and total execution time. Figure 1 shows the patterns and the velocity profiles for one of their executions.

The paths extracted from the motion data were processed to obtain the trajectory in three different ways, depending on the time law assigned to the path: the one derived from the online data, the one provided by the STOTPAC algorithm, and a constant tangential velocity profile. These trajectories, hereinafter denoted as *Human*, *STOTPAC* and *Constant*, respectively, were used as follows:

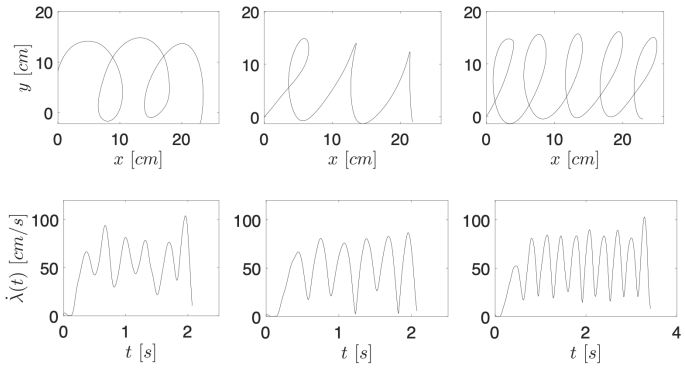


Fig. 1. The representation of the trajectories resulting from one execution of each pattern by one subject: the paths (top panel) and the velocity profiles (bottom panel).

- we compared the *Human* velocity profile with the *STOTPAC* one for evaluating the accuracy of the latter profile to estimate the former one;
- we compared the *Human* with the *Constant* one for evaluating the accuracy of a non-human profile to estimate the *Human* one.

The last comparison was designed to provide the upper bound for the first, as will be explained in the next subsection.

B. Performance metrics and results

For comparing the velocity profiles, we used the Dynamic Time Warping (DTW), an algorithm widely adopted in signal processing to measure the similarity between two temporal sequences, which may vary in speed. The algorithm provides as output a "distance" between the sequences, hereinafter referred to as *DTW-distance*, and thus high values indicate low similarity between the trajectories under comparison. It must be noted that although DTW measures a distance-like quantity between two given sequences, it does not guarantee that the triangle inequality holds.

We performed two experiments. In the first one, for each pattern and each writer, we compared the *Human* velocity profiles with those provided by the *STOTPAC* and the *Constant* algorithms. Because of space limitation, Table I reports the results for the executions of the second pattern of Figure 1 by a single subject. The data show that STOTPAC provides a much better approximation of the *Human* profile than the Constant algorithm. The conclusion is supported by a nonparametric t-test with Welch's correction ($\alpha = 0.95$, $p\text{-}value < 0.05$) that confirms a statistically significant difference between the mean and the variance of the DTW-distances. As similar data were obtained independently of the pattern and the subject, we conclude that the STOTPAC algorithm provides better estimations of *Human* handwriting velocity profiles than the Constant algorithm. Thus, the second experiment is restricted to the former algorithm.

TABLE I
DTW-DISTANCE FROM THE FIRST EXPERIMENT AND THEIR MEAN AND STANDARD DEVIATION.
FIRST ROW: *Human* VS *STOTPAC*. SECOND ROW: *Human* VS *Constant*.

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**mean**	**std**
8475.62	8993.42	8169.46	12330.30	9183.27	9381.52	8226.97	10990.19	12149.32	12385.75	10028.58	1751.03
57371.67	52084.80	51714.12	43747.70	50465.34	49248.68	59964.58	56118.56	53530.91	51825.19	52607.15	4546.72

In the second experiment, we compared, for each pattern, the *Human* profile of a handwriting sample with the *Human* and *STOTPAC* profiles of the remaining 9 samples written by the same subject. In fact, because the DTW-distance does not guarantee that the triangular inequality holds, the comparison is carried out by assuming each *Human* execution as reference and evaluating the DTW with the remaining 9 executions, resulting in 18 comparisons for each reference, thus totaling 180 comparisons for each subject and pattern. For the same reason as before, Table II reports the DTW-distance of the 9 comparisons on the same pattern and for the same subject of Table I when the reference is the first execution, with its mean and standard deviation. Moreover, the statistical analysis performed as in the previous experiment shows that while the difference in the mean among the two distributions is statistically significant, the difference between the variances is not. The mean and standard deviation in the first row of Table II, moreover, show that all the DTW-values reported in the first row of Table I are within the interval ($\mu \pm 3\sigma$) of the DTW-distance among the *Human* profiles. Both the previous observations hold for each subject, independently of the reference execution and the pattern to be executed. Thus, we conclude that the STOTPAC algorithm generates velocity profiles that resemble the *Human* profiles and preserve the intra-writer variability.

TABLE II
DTW-DISTANCE FROM THE SECOND EXPERIMENT WHEN THE FIRST EXECUTION IS USED AS THE REFERENCE IN THE COMPARISON. THE FIRST ROW INCLUDES THE DTW-DISTANCES BETWEEN THE *Human* PROFILE OF THE REFERENCE AND THE *Human profile* OF ANOTHER EXECUTION OF THE SAME PATTERN AND THE SAME WRITER. THE SECOND ROW INCLUDES THE DTW-DISTANCES BETWEEN THE *Human* PROFILE OF THE REFERENCE AND THE *STOTPAC* PROFILE COMPUTED FOR ANOTHER EXECUTION OF THE SAME PATTERN AND THE SAME WRITER.

	1 vs 2	**1 vs 3**	**1 vs 4**	**1 vs 5**	**1 vs 6**	**1 vs 7**	**1 vs 8**	**1 vs 9**	**1 vs 10**	**mean**	**std**
Human	6652.12	11217.24	6785.65	4629.97	7127.64	4927.75	7070.80	7819.77	8382.47	7179.27	1942.25
STOTPAC	5097.36	6538.47	5862.75	4359.95	6405.90	4680.16	5007.50	4494.17	4667.56	5234.87	827.58

C. Discussion

The experimental results and the conclusions we have drawn were obtained assuming that the *Human* trajectory was available, but this is not the case in the forensic examination. Applying the proposed procedure to a real-world case requires recovering the path of pen-tip trajectories from images of handwriting samples using one of the several approaches proposed in the literature [5]. Moreover, the forensic examiner could assume as *patterns* the n-grams common to both the genuine and the questioned documents (e.g. bigrams, trigrams, up to whole words), and as samples the *executions* of the corresponding patterns. Within this setting, and taking into account the previous evidence that the *STOTPAC* velocity profiles are a good approximation of the unknown velocity profile and preserve the intra-writer variability, the comparison in a real-world case can be implemented as follows:

1) for each pattern:
   for each sample, get the representation of the path by a thin line that incorporates the writing order of the ink trace;
   compute the *STOTPAC* velocity profiles associated with the paths of the *Genuine* samples;
   compute the *STOTPAC* velocity profiles associated with the paths of the *Questioned* samples;
2) compute the intra-writer variability of both the *Genuine* and the *Questioned* samples as in the experiment 2.

At the end of this procedure, thus, the statistical analysis will give the examiner evidence on whether or not the *Genuine* and the *Questioned* samples of each pattern were executed by the same subject, and the p-value can be used to weigh the extent to which each pattern contributes to the final decision, as well as to guide the next steps of the examination procedure. Preliminary results, not reported here for the sake of space, obtained on a small data set of handwriting samples simultaneously collected in both online and offline, have shown promising results along this direction for writer identification.

IV. CONCLUSION

Our work aims at providing a better characterization of offline samples for forensic handwriting examination. For the purpose, we have proposed a variant of the TOTP algorithm, largely adopted in planning robot motion along an assigned path, to associate a velocity profile to the handwriting ink trace, with the ultimate aim of comparing a genuine specimen versus a questioned one.

The results of the experiments we have designed and performed show that enriching the path of the ink with the velocity profile provided by our algorithm is a good approximation of the human one and preserves the intra-writer variability. Eventually, based on these findings, we have outlined a procedure for implementing a thoughtful and quantitative comparison between the handwriting under examination, and show how the results may provide robust evidence to the examiner for achieving the final decision.

Our future work will focus on building a large data set of handwriting samples, simultaneously collected in online and offline modes, testing the robustness of our conclusion, and comparing the results achieved by using the same automatic writer identification system on both the online and the offline samples, when the latter are assigned the STOTPAC velocity profiles.

REFERENCES

[1] Jaya Paul; Kalpita Dutta; Anasua Sarkar; Nibaran Das; Kaushik Roy. A survey on different feature extraction methods for writer identification and verification. *Int, Journal of Applied Patter Recognition (IJAPR)*, 7(2):122–144, 2023.
[2] Donato Impedovo and Giuseppe Pirlo. Automatic signature verification: The state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 38(5):609–635, 2008.
[3] J.-J.E. Slotine and H.S. Yang. Improving the efficiency of time-optimal path-following algorithms. *IEEE Transactions on Robotics and Automation*, 5:118–124, 1989.
[4] Hands Leo Teulings. MovAlyzeR. version 6.1, 2019.
[5] Rosa Senatore, Adolfo Santoro, Antonio Parziale, and Angelo Marcelli. A biologically inspired approach for recovering the trajectory of offline handwriting. *Memetic Computing*, 15(3):355–375, 2023.

Evaluating Word Embedding Models for Authorship Attribution

Hiba Fallatah
CENPARMI Lab, Concordia University
Montreal, Quebec, Canada
h_fall@encs.concordia.ca

Ching Suen
CENPARMI Lab, Concordia University
Montreal, Quebec, Canada
suen@cse.concordia.ca

***Abstract*— This study compares the performance of six word embedding models—Word2Vec, GloVe, ELMo, BERT, RoBERTa, and XLM-RoBERTa—for authorship attribution in chat logs. Experiments were conducted with varying numbers of authors (2, 5, 10, and 20) to evaluate accuracy and F1-score. The results show that ELMo consistently outperforms the other models, particularly in more complex scenarios with a higher number of authors. Transformer-based models (BERT, RoBERTa, and XLM-RoBERTa) also perform well, especially in low-author settings. Word2Vec outperforms GloVe in most cases, particularly when the number of authors is small. The study highlights the importance of context-sensitive embeddings like ELMo and BERT in identifying authors and calls for further research to improve performance in high-author settings.**

Keywords*— *Word embedding, Author attribution, Graph neural network

I. Introduction

Word embeddings represent words as dense, fixed-length vectors that capture semantic meaning and contextual relationships [1], [2]. These embeddings have significantly advanced natural language processing (NLP) and machine learning applications, improving tasks such as text classification, sentiment analysis, and machine translation [3]. One important application of word embeddings is authorship attribution, which involves identifying the author of a given text based on linguistic and stylistic features. Since word embeddings can capture subtle syntactic and semantic variations in writing style, they are well-suited for this task.

Authorship attribution is essential in various domains, including literary analysis, forensic linguistics, and plagiarism detection [4]. The process relies on identifying unique patterns in an author's writing style, such as word choice, sentence structure, and vocabulary usage. With the advancement of NLP techniques, word embeddings have evolved from static models like Word2Vec to contextual models such as BERT, which better capture the deeper meaning of words in different contexts. This evolution has led to notable improvements in NLP tasks, including authorship attribution.

This study evaluates the effectiveness of several word embedding models for authorship attribution in chat logs. It includes static embeddings such as Word2Vec and GloVe, which capture semantic relationships based on word co-occurrence, and contextual embeddings such as ELMo, BERT, RoBERTa, and XLM-RoBERTa, which generate dynamic representations based on sentence context. Specifically, the study investigates which models best capture distinguishing features of an author's writing style, particularly in challenging scenarios involving multiple authors and limited text samples. The findings provide insights into each model's strengths and limitations, highlighting the advantages of transformer-based models in handling stylistic nuances. This work also underscores the broader challenges of applying word embeddings to authorship attribution in dynamic and informal text environments.

Beyond authorship attribution, which is the focus of this study, Authorship analysis includes several related tasks, including authorship verification, writing style comparison, and the detection of AI-assisted or machine-generated content. Authorship verification determines whether two texts were written by the same person without assuming a fixed candidate set [5]. This open-world formulation is particularly relevant in contexts such as plagiarism detection and online identity validation. Writing style comparison analyzes stylistic similarities or differences across texts, which is valuable for identifying ghostwriting, collaborative authorship, or tracking the stylistic evolution of an author [6]. Meanwhile, the detection of AI-generated text has gained increasing importance with the rise of generative models like ChatGPT, requiring techniques that can distinguish between human and machine-produced content and whether a text was produced in part or entirely by an automated system [7], [8]. While each of these tasks presents unique challenges, they all share an underlying reliance on stylistic features and linguistic cues, underscoring the broader significance and complexity of authorship analysis in the era of advanced language models.

II. Overview

This section establishes the theoretical foundation for the study by discussing word embedding techniques grouped into: static embeddings (Word2Vec, GloVe), deep contextualized embeddings (ELMo, BERT), and advanced transformer-based models (RoBERTa, XLM-RoBERTa). These models are introduced and analyzed in terms of their relevance to authorship attribution tasks.

A. Techniques for Representing Words in Vector Space

Word embedding methods can be categorized into four main types [9]:

a) Frequency-based Methods: Determine word significance based on occurrence and co-occurrence. While simple and interpretable, these methods lack contextual understanding and struggle with polysemy [10]. (e.g., Bag of Words, TF-IDF)

b) Static-based Methods: Represent words as fixed, dense vectors without considering context. They capture semantic and syntactic relationships [11], but fail to account for context-dependent meanings of words and have a limited ability to handle Out-of-Vocabulary (OOV) and uncommon words [9]. (e.g., Word2Vec, GloVe)

c) Contextual-based Methods: Generate different embeddings for words depending on their context, effectively handling polysemy and out-of-vocabulary words. However, these models require significant computational resources, and if not utilized appropriately, it could lead to overfitting the training set [12]. (e.g., BERT, ELMo)

d) Sentiment-aware Methods: Incorporate emotional tone into word embeddings, making them useful for sentiment analysis but not necessarily for all NLP tasks. Nevertheless, sentiment-aware techniques are capable of interpreting linguistic nuances connected to sentiments [13].

B. Popular Word Embedding Models

- **Word2Vec** [14] Learns word representations by predicting context words. It uses two architectures: Continuous Bag of Words (CBOW) and Skip-gram. Word2Vec is effective in capturing word similarity and local context, making it useful for authorship attribution.
- **GloVe** [15] Captures global word co-occurrence statistics, creating embeddings that reflect broader semantic relationships. While it effectively models word associations, it may miss subtle stylistic differences crucial for authorship attribution.
- **ELMo** [16] Produces word embeddings from a deep, bidirectional LSTM trained on a large text corpus. Unlike static embeddings, which assign a single vector representation to each word, ELMo dynamically adjusts word embeddings based on surrounding context. This allows it to capture contextual variation in word usage and can represent subtle stylistic differences across authors more effectively than Word2Vec or GloVe.
- **BERT** [17] Generates context-sensitive word embeddings using a transformer-based approach. By capturing bidirectional context, BERT is well-suited for distinguishing writing styles, making it the most promising model for authorship attribution.
- **RoBERTa** [18] Is an improved version of BERT that removes the next-sentence prediction objective, allowing it to focus entirely on masked language modelling. It is trained with larger mini-batches, more data, and longer training durations, leading to better performance on various NLP benchmarks.
- **XLM-RoBERTa** [19] Is a multilingual extension of RoBERTa, pre-trained on 100 languages using 2.5 terabytes of CommonCrawl data. Unlike standard RoBERTa, XLM-RoBERTa is optimized for cross-lingual tasks, making it highly effective for multilingual authorship attribution and language understanding.

III. Experimental Setup

This section outlines the dataset, preprocessing steps, and model training process used to evaluate a range of word embedding models—including static (Word2Vec, GloVe), deep contextualized (ELMo, BERT), and advanced transformer-based embeddings (RoBERTa, XLM-RoBERTa)—for authorship attribution in chat logs.

A. Dataset Description

The study uses the PAN 2012 dataset [20], which contains chat logs from four sources: Omegle, IRC, Krijn, and Perverted-Justice. The dataset includes 221,871 conversations with 316,438 participants, of which only 396 are identified as predators. To prepare the data, each document is structured to contain all chat logs from a single author within a specific conversation. To reduce noise, numbers, timestamps, and URLs are replaced with placeholders (e.g., NUM, TIM, URL).

B. Model Training Process

The study utilizes a Graph Neural Network (GNN), specifically a Graph Attention Network (GAT), to train the authorship attribution models. Each document is represented as a graph, where words act as nodes and relationships between words form edges. Two types of edges are defined: dependency-based edges, which capture syntactic relationships between words, and sequential-based edges, which encode the order of words in a sentence. The input features for each node are derived from pretrained word embeddings obtained from Word2Vec, GloVe, BERT, RoBERTa, XLM-RoBERTa, or ELMo. The weight of each edge is determined by the frequency of word co-occurrence within a document, reflecting the strength of connections between words.

During training, the GNN processes these graphs and applies attention mechanisms to aggregate information from neighbouring words, refining the embeddings to better distinguish between authors. The model learns to recognize stylistic patterns and linguistic structures unique to each writer. The final output of the GNN is used to classify the document's author.

This graph-based representation enhances authorship attribution by leveraging both the semantic meaning of words and the structural characteristics of the text.

IV. Results

To evaluate the effectiveness of word embedding models for authorship attribution, we compared the six embedding-based approaches—Word2Vec, GloVe, ELMo, BERT, RoBERTa, and XLM-RoBERTa—using the PAN 2012 dataset. The results are summarized in Table I.

TABLE I. Performance comparison of word embedding models for authorship attribution

Embedding Model	2 Authors		5 Authors		10 Authors		20 Authors	
	Acc.	**F1**	**Acc.**	**F1**	**Acc.**	**F1**	**Acc.**	**F1**
RoBERTa	98.09	98.08	68.73	67.90	50.94	48.52	41.90	36.14
XLM-RoBERTa	99.28	99.27	69.82	69.48	54.70	51.63	43.17	37.12
BERT	98.60	98.59	71.87	69.33	53.22	49.06	44.64	38.98
ELMo	**99.90**	99.99	**74.86**	72.43	**62.29**	60.85	**52.17**	48.62
GloVe	89.89	89.06	54.20	44.61	39.71	30.02	33.25	23.02
Word2Vec	94.41	94.19	61.24	55.64	45.35	39.97	38.53	31.25

To test the efficiency of using word embedding models for authorship attribution, we compared it with two traditional approaches: [21] and [22]. The first approach, based on SVM, uses a combination of lexical, syntactic, and content-specific features to identify authors. The second approach relies on character 4-grams, computing authorship similarity using the Cosine distance.

Results from the PAN 2012 dataset show that ELMo significantly outperforms both traditional methods. In the 2-author attribution task, ELMo achieves the highest accuracy (99.90%), with a recall and precision of 99.90%, demonstrating its superior ability to capture writing style compared to SVM and Cosine Similarity. Table II summarizes the comparison of our preliminary results with the existing methods over the PAN 2012 dataset to identify the unknown text between the two authors.

TABLE II. Performance comparison of using word embedding model (ELMo) vs. existing methods

classification algorithms	**Accuracy**	**Precision**	**Recall**
Cosine similarity [22]	36.93 %	57.07	49.94
SVM [21]	79.78 %	74.91	69.76
ELMo	**99.90 %**	99.90	99.90

A. Performance Analysis

As shown in Fig. 1, ELMo achieves the highest performance across all tasks, although the margin over transformer-based models such as BERT, RoBERTa, and XLM-RoBERTa is relatively small in simpler scenarios. The advantage of ELMo comes from its use of deep, contextualized word representations generated via a bidirectional LSTM architecture. Unlike transformers, which process tokens in parallel with position encodings, ELMo processes text sequentially in both forward and backward directions. This enables it to effectively capture local syntactic dependencies and stylistic cues that are especially relevant in authorship attribution. In addition, ELMo utilizes character-level convolutional layers to generate word representations, allowing it to effectively handle misspellings and out-of-vocabulary words. By building representations from character sequences rather than relying solely on word-level embeddings, ELMo can infer meaning even from rare or unseen words. This fine-grained modelling is particularly beneficial for detecting subtle stylistic markers. While transformer models generally outperform ELMo in broader NLP tasks due to their capacity for modelling long-range dependencies, ELMo appears especially well-suited for capturing the short-range, style-specific signals characteristic of informal texts like chat data.

Although ELMo's performance is only slightly better in settings with fewer authors, the gap becomes more noticeable as the number of authors increases. This suggests that ELMo scales more effectively in capturing distinct writing styles as the attribution task becomes more complex.

In comparison, BERT, RoBERTa, and XLM-RoBERTa exhibit very similar performance, all clearly outperforming static embeddings like Word2Vec and GloVe. These transformer-based models use self-attention mechanisms to model long-range dependencies and are pretrained using masked language modelling (MLM), which enables rich, bidirectional context learning. BERT incorporates both MLM and next sentence prediction (NSP), which helps it model not only word-level context but also relationships between consecutive sentences. RoBERTa refines BERT by removing the NSP objective, increasing training data and batch sizes, and employing dynamic masking strategies. XLM-RoBERTa extends RoBERTa's architecture to multilingual

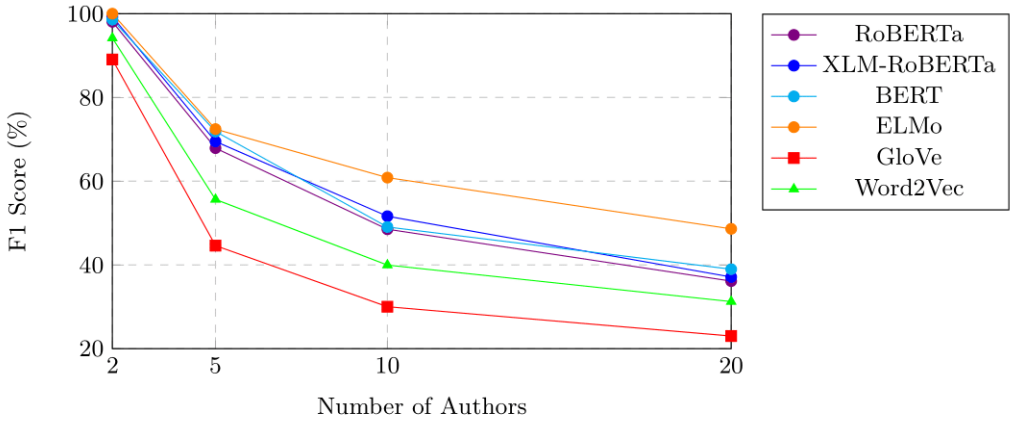


Fig. 1. Performance comparison of word embedding models for authorship attribution (F1 score vs. number of authors).

data, improving generalization across languages but potentially reducing its effectiveness in monolingual English tasks. Despite these architectural variations, the similar results across BERT, RoBERTa, and XLM-RoBERTa indicate a performance plateau on this dataset. This may reflect shared pretraining strategies and a common limitation in adapting to domain-specific traits of chat-based text.

Among the static models, Word2Vec consistently outperforms GloVe, particularly in both simpler and more complex tasks. Word2Vec, especially its Skip-gram architecture, emphasizes local context by predicting neighbouring words for a given target word, making it more effective at capturing localized stylistic patterns typical of specific authors. Word2Vec also handles rare words better, as it learns from individual word-context pairs, which is advantageous for identifying distinctive word usage.

In contrast, GloVe captures global co-occurrence statistics across the entire corpus, which is helpful for understanding general semantic relationships but may overlook fine-grained stylistic cues necessary for authorship discrimination. As a result, it underperforms compared to both Word2Vec and contextual models in tasks involving more than two authors.

V. Conclusion

The study evaluates the effectiveness of six word embedding models for authorship attribution in chat logs. The results show that ELMo consistently achieves the highest performance, particularly as the number of authors increases, due to its deep contextualized embeddings that effectively capture nuanced syntactic and semantic patterns. Among transformer-based models, BERT also performs strongly, outperforming static embeddings like Word2Vec and GloVe thanks to its ability to capture fine-grained, context-dependent stylistic features. Word2Vec performs better than GloVe, likely due to its strength in modeling local word context, which is critical for distinguishing individual writing styles. RoBERTa and XLM-RoBERTa offer competitive performance but perform marginally below ELMo and BERT in this task, possibly due to tokenization effects or domain mismatch with chat data. As the number of authors increases, all models exhibit a decline in performance, highlighting the challenge of high-author attribution scenarios and the need for further research in this area. While contextual models offer superior accuracy, their computational cost remains a constraint. Future work could investigate integrating stylistic, syntactic, and discourse-level features or enhancing graph neural network architectures to better model authorship traits. Although this study focuses on closed-set authorship attribution, extending to tasks like authorship verification or AI-generated content detection presents a logical progression. These tasks differ in formulation but share underlying stylistic cues that embedding models can learn. Exploring how transformer-based embeddings generalize across such tasks—and whether training on mixed human-AI datasets boosts robustness—offers promising directions for future research.

References

[1] F. Almeida and G. Xexéo, “Word Embeddings: A Survey,” May 2023. [Online]. Available: http://arxiv.org/abs/1901.09069
[2] E. Sezerer and S. Tekir, “A Survey On Neural Word Embeddings,” Oct. 2021. [Online]. Available: http://arxiv.org/abs/2110.01804
[3] D. S. Asudani, N. K. Nagwani, and P. Singh, “Impact of word embedding models on text analytics in deep learning environment: a review,” *Artificial Intelligence Review*, vol. 56, no. 9, pp. 10 345–10 425, Sep. 2023. [Online]. Available: https://link.springer.com/10.1007/s10462-023-10419-1
[4] E. Stamatatos, “A survey of modern authorship attribution methods,” *Journal of the American Society for Information Science and Technology*, vol. 60, no. 3, pp. 538–556, Mar. 2009. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/asi.21001
[5] C.-Y. Hung, Z. Hu, Y. Hu, and R. K.-W. Lee, “Who Wrote it and Why? Prompting Large-Language Models for Authorship Verification,” Oct. 2023, arXiv:2310.08123 [cs]. [Online]. Available: http://arxiv.org/abs/2310.08123
[6] A. Hashemi and W. Shi, “A Survey on Writing Style Change Detection: Current Literature and Future Directions,” *Machine Intelligence Research*, Apr. 2025. [Online]. Available: https://doi.org/10.1007/s11633-025-1544-6

[7] J. Weerasinghe, O. Seepersaud, G. Smothers, J. Jose, and R. Greenstadt, "Be Sure to Use the Same Writing Style: Applying Authorship Verification on Large-Language-Model-Generated Texts," *Applied Sciences*, vol. 15, no. 5, p. 2467, Jan. 2025, number: 5 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2076-3417/15/5/2467

[8] A. Richburg, C. Bao, and M. Carpuat, "Automatic Authorship Analysis in Human-AI Collaborative Writing," in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, and N. Xue, Eds. Torino, Italia: ELRA and ICCL, May 2024, pp. 1845–1855. [Online]. Available: https://aclanthology.org/2024.lrec-main.165/

[9] S. J. Johnson, M. R. Murty, and I. Navakanth, "A detailed review on word embedding techniques with emphasis on word2vec," *Multimedia Tools and Applications*, vol. 83, no. 13, pp. 37 979–38 007, Oct. 2023. [Online]. Available: https://link.springer.com/10.1007/s11042-023-17007-z

[10] K. Zhou, K. Ethayarajh, and D. Jurafsky, "Frequency-based Distortions in Contextualized Word Embeddings," Apr. 2021. [Online]. Available: http://arxiv.org/abs/2104.08465

[11] P. Gupta and M. Jaggi, "Obtaining Better Static Word Embeddings Using Contextual Embedding Models," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, C. Zong, F. Xia, W. Li, and R. Navigli, Eds. Online: Association for Computational Linguistics, Aug. 2021, pp. 5241–5253. [Online]. Available: https://aclanthology.org/2021.acl-long.408

[12] Q. Liu, M. J. Kusner, and P. Blunsom, "A Survey on Contextual Embeddings," Apr. 2020. [Online]. Available: http://arxiv.org/abs/2003.07278

[13] X. Mao, S. Chang, J. Shi, F. Li, and R. Shi, "Sentiment-Aware Word Embedding for Emotion Classification," *Applied Sciences*, vol. 9, no. 7, p. 1334, Jan. 2019. [Online]. Available: https://www.mdpi.com/2076-3417/9/7/1334

[14] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," Sep. 2013. [Online]. Available: http://arxiv.org/abs/1301.3781

[15] J. Pennington, R. Socher, and C. Manning, "Glove: Global Vectors for Word Representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, 2014, pp. 1532–1543.[Online]. Available: http://aclweb.org/anthology/D14-1162

[16] S. Ilić, E. Marrese-Taylor, J. A. Balazs, and Y. Matsuo, "Deep contextualized word representations for detecting sarcasm and irony," Sep. 2018, arXiv:1809.09795 [cs]. [Online]. Available: http://arxiv.org/abs/1809.09795

[17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," May 2019. [Online]. Available: http://arxiv.org/abs/1810.04805

[18] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," Jul. 2019, arXiv:1907.11692 [cs]. [Online]. Available: http://arxiv.org/abs/1907.11692

[19] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised Cross-lingual Representation Learning at Scale," Apr. 2020, arXiv:1911.02116 [cs]. [Online]. Available: http://arxiv.org/abs/1911.02116

[20] G. Inches and F. Crestani, "Overview of the International Sexual Predator Identification Competition at PAN-2012," Jan. 2012.

[21] F. Amuchi, A. Al-Nemrat, M. Alazab, and R. Layton, "Identifying Cyber Predators through Forensic Authorship Analysis of Chat Logs," in *2012 Third Cybercrime and Trustworthy Computing Workshop*, Oct. 2012, pp. 28–37. [Online]. Available: https://ieeexplore.ieee.org/document/6498425

[22] N. M. Sharon Belvisi, N. Muhammad, and F. Alonso-Fernandez, "Forensic Authorship Analysis of Microblogging Texts Using N-Grams and Stylometric Features," in *2020 8th International Workshop on Biometrics and Forensics (IWBF)*. Porto, Portugal: IEEE, Apr. 2020, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/9107953/

Towards a Gold Standard for AI Facial Landmarks Estimation: Constructing FAIRSET, a Balanced and Inclusive Landmark Database

Nikola Zelovic[1,3], Ian-Mathieu Joly[1,3], Eléonor Riesco[2,3], Karina Lebel[* 1,3]

[1]Faculty of Engineering, Université de Sherbrooke, Sherbrooke, QC, Canada
[2]Faculty of Physical Activity Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada
[3]Research Center on Aging, CIUSSS de l'Estrie - CHUS, Sherbrooke, QC, Canada

[*]karina.lebel@usherbrooke.ca

***Abstract*— Facial landmarks detection is used for many applications, but the inherent biases of the models are not well described in the literature, nor understood. This study aims to build FAIRSET: an inclusive and balanced dataset of faces with face landmark annotations for the purpose of evaluating said models.**

Keywords*— *Face Landmarks, Facial Analysis, Inclusive Dataset, Deep Learning Bias, AI Algorithms Evaluation.

I. Introduction

Facial landmarks detection (FLD) can be a useful tool for recognizing and targeting facial features, in images or videos, to be used in a broader applications such as expression recognition [1] and heart rate estimation [2]. Several deep learning models and facial landmark detection toolboxes can accomplish this task, such as MediaPipe face mesh [3], 3DDFAv3 [4], and OpenFace 2.2 [5]. However, little information is available on the possible biases they may contain affecting their detection performance on specific demographics. Multiple studies evaluated the different biases for face detection tasks (which are often the first step for FLD) and found risks of biases for females, older populations and darker skin tones [6-7]. Similar biases were also found for other AI tasks such as gender classification [8]. Despite that, to our knowledge, no bias analysis was conducted specifically for face landmarks detection. Yet, face landmarks detection accuracy may affect subsequent uses such as heart rate estimation via RGB cameras [9]. To better understand the limitations of face landmarks detection models and to ensure inclusive usage, this study aims to create the Face Annotations, Inclusivity & Representation dataSET (FAIRSET). A balanced database of face landmarks that includes annotations about the sex, age and skin tone of the pictured individuals. Databases containing a fair and balanced collection of people's faces exist, however, they are more targeted towards face recognition and are not suitable for face landmarks evaluation. The proposed FAIRSET database ensures that all faces can be recognized by mainstream FLD models and toolboxes, while also providing a balanced set of annotations across different demographics. This dataset can be used to assess and compare the precision of FLD tools on both static images and video (by averaging the inter-frame precision), ensuring a high-level of confidence in the targetting of face regions or in the tracking of face movements.

II. Database Creation Principles

The proposed database was developed in four main steps: (1) *image selection* based on a combination of sex, age, and skin tone from existing databases; (2) *image annotation* using Amazon Mechanical Turk; (3) *annotations verification*; (4) *evaluation* of accepted annotations across different annotators and comparison to the referenced labels for available attributes.

A. FAIRSET Image Selection

Amazon Alexa WiderFace (AWF) subset [10] and Dad-3DHeads [11] databases were selected as starting points, because even though they present an unbalanced ratio of annotated people across sex, age and skin tone, they do contain annotations about demographics attributes. Table I presents the different information available in each database's annotations.

To select the FAIRSET images, a series of specific steps was taken. First, a preliminary analysis was conducted by running popular FLD models like MediaPipe [3], 3DDFAv3 [4], and OpenFace 2.2 [5] on the Alexa WiderFace database to determine the minimum bounding box size required for the algorithms to work. This analysis revealed that MediaPipe and 3DDFAv3 had difficulty detecting faces with bounding boxes under a width of approximately $200px$ and a height of approximately $300px$. Based on these results, only images with a maximum of two faces and with bounding boxes of 200x300px or larger were selected for the creation of FAIRSET. A target of 1000 total images, evenly distributed across 48 subgroups, was chosen. These subgroups were based on the combination of criteria of interest, namely sex (male/female), age group (young adult, adult, middle-aged, senior) and skin tone (I to VI on the Fitzpatrick scale). Images were first selected from the AWF database and

TABLE I: Annotated Attributes Across Databases

	Alexa WiderFace	Dad-3DHeads
Face landmarks	✗	✔*
Bounding box	✔	✔
Sex	✔	✔
Age groups	• 18 and under • Young Adult • Adult • Middle-aged • Senior	• Child • Young • Middle-aged • Senior
Skin tone	✔**	✗
Head pose	✗	✔

* Vertices from the 3D model, instead of face landmarks targeted from face features.
** Skin tones based on the Fitzpatrick scale

TABLE II: Distribution of all selected imaged, by sex and age

Gender	Age Group	Wider Face	Dad-3DHeads	Re-moved	Total
Female	**Child**	0	50	9	41
	Young Adult	66	50	7	109
	Adult	69	0	0	69
	Middle-aged	4	100	7	97
	Senior	7	100	15	92
Subtotal		146	300	38	**408**
Male	**Child**	0	50	8	42
	Young Adult	64	50	4	110
	Adult	93	0	0	93
	Middle-aged	25	75	10	90
	Senior	16	100	18	98
Subtotal		198	275	40	**433**
Total		344	497	78	**841**

then augmented with images from the DAD-3DHeads database to ensure balanced representation, aiming for 100 images per subgroup, except for the *child* age group for which 50 images were targeted. Finally, black and white images were manually removed. Table II present the totals of images taken from each database and how they intersect across different sexes, age groups and skin tones.

B. Annotation of face landmarks and skin tone by Mechanical Turk workers

Mechanical Turk (MTurk) is a crowdsourcing platform that allows, among other things, to submit tasks for keypoint annotation. An MTurk template was adapted to automate some quality checks and to enable collection of additional information, such as skin tone. Every image was submitted to at least two different annotators and 34 keypoints were initially targeted for the creation of the FAIRSET database, with two additional keypoints (the clavicles) subsequently added, for a total of 36 keypoints. Some keypoints were chosen because they are universal across FLD models (*e.g.*, chin, nose tip, corners of eyes) and the others were selected to allow analysis of face movements (*e.g.*, points on the face contour, eyebrows, mouth). The shoulders and the sternum were added to facilitate orientation estimations. The approximate locations of the different keypoints are shown in figure 1.

C. Two-steps validation of the annotated face keypoints

Received annotations were first automatically validated to ensure no duplicated keypoints were present and that the number of annotated skin tones matched the number of people present in the image. Then, every annotated image was manually verified to make sure that there were no egregious placements of keypoints and that the skin tone made approximate sense. Annotations that had a few small mistakes were still accepted, with the validator manually fixing the issue. Rejected images were republished on MTurk. As a final step, accepted images underwent a "quality enhancement" pass. During this step, a validator regrouped keypoints associated with the same person in images containing multiple individuals. Additionally, labels were corrected if the annotator selected the wrong one and the associated keypoint was not in the labelled location.

D. Evaluation and Quality Assurance

To ensure accuracy of the image annotations within the FAIRSET database, the final stage of its creation involved a review of potential ambiguities related to the various criteria. An evaluation of the accuracy and reliability of the initial annotations, and the generation of the final annotations, were executed as well. The AWF *Adult* and the DAD-3DHeads *Middle-aged* age groups were combined, since visual inspection of their images identified an overlap. To evaluate accuracy and reliability of the annotations, the interocular normalized mean error (NME) metric was used. This metric refers to the Euclidean error of a point in respect to its reference, normalized with the distance between the inner-corner of the eyes. This NME allows to characterize the difference between annotators over a same face and to assess reliability of keypoint annotations. Statistical analyses were conducted to appraise possible biases at the annotator level. The quality of the skin tone annotation was also evaluated on the subgroup of images sourced from Alexa WiderFace. Final labels were then generated by combining multiple annotations of the same face. The 2D keypoint coordinates were averaged, and skin tones were merged by rouding up the average value. If there was a difference of two or more skin tones between raters, a human validator reviewed and carefully selected the appropriate label. The finalized annotations include sex and age group from the original databases' labels.

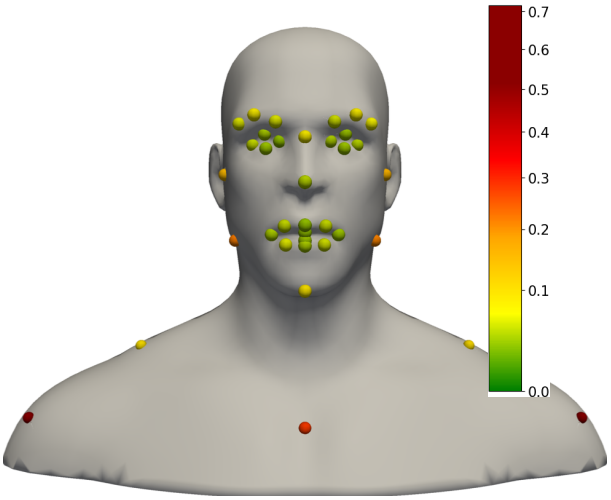


Fig. 1: Average inter-annotator error per keypoint

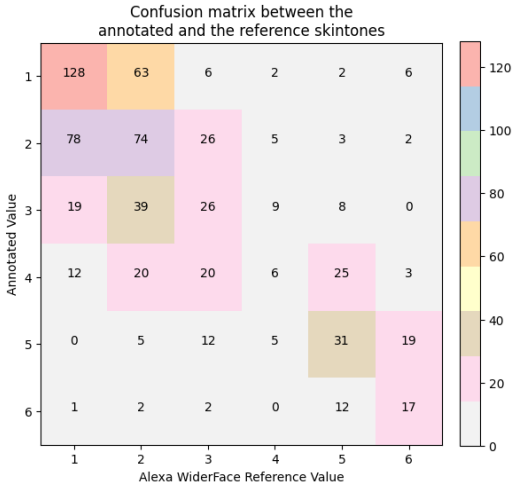


Fig. 2: Confusion matrix between MTurk skin tone annotations and AWF reference labels

III. Results

A total of 3446 annotated images were received, of which 1526 were accepted for a total of 1612 annotated individuals. Among these, 1589 had a reference label for sex and age group. The distribution of received annotations is balanced across these two attributes, with the smallest adult group being *Female Adult* with 8.2% of all annotations, and the largest adult group (*Female Young Adult*) receiving 13.1%. The received skin annotations are compared to the Alexa WiderFace reference in figure 2. The presented confusion matrix shows a few mismatched skin tones, indicating that the annotator did not try to properly annotate the skin tone or that the reference label is mistaken. The resulting error in tones compared to the reference values has a median of 1.00 and a mean error of 1.85 ± 1.59 between the average of the workers and the reference. However, the consistency across workers annotating a same face is of higher quality with a mean error of 0.70 ± 0.76 for all faces across the dataset with more than one annotation, and of 0.76 ± 0.75 for Alexa WiderFace images. Since the estimated skin tone is subjective and can also be skewed by lighting and background, it is considered that the received annotations can be relied upon for the final dataset. The extremes shown in the matrix were manually corrected in a second pass to produce the final dataset labels, presented in figure 4. The global NME average between annotators for a same face is 0.11 ± 0.29. However, the positioning of the shoulders and of the sternum posed a significant challenge for the annotators, increasing the global error. Removing these three keypoints from the analysis reduces the global NME to 0.07 ± 0.10. Figure 1 depicts the 36 keypoints on an avatar, with different colored degrees of error. Multiple one-way ANOVA analyses on the NME values were conducted for each keypoint across sex, age and skin tone. These analyses showed no notable statistical difference or errors between annotators, except for a few keypoints between males and females. However, the keypoints that stand out do not show symmetry on the face, casting doubt on true annotator bias. Finally, annotators demonstrated linear correlation in their keypoint annotations with a Pearson's correlation coefficient above 0.95 for all keypoints except for shoulders.

The final FAIRSET database consists of 810 images, featuring 866 faces. The labels encompass sex, age group, skin tone, 2D coordinates for 36 keypoints, and head orientation when available. The sex and age group labels were taken from Alexa WiderFace and DAD-3DHeads labels, after merging adults and middle-aged groups. Head pose information was extracted from their original DAD-3DHeads labels, when appropriate. Figure 3 offers an overview of the final distribution of the labels. The confusion matrix (figure 4) also shows the corrected skin tones for AWF images, with a mean skin tone difference of 0.79 ± 0.83 from the reference annotation, and with 86% of the final skin tones falling within one tone of the reference.

IV. Discussion

The final FAIRSET annotations are approximately balanced for sex and age for adult faces, with the smallest group being *Female Seniors* (89 annotated faces). Evaluating biases caused by skin tone might be more difficult for groups other than male adults, due to the limited number of images with darker skin tones, as seen in figure 3. Addressing this shortcoming by sourcing additional images should be a priority for a future iteration of the database. The final distribution FAIRSET is 47.2% females, 50.5% males and 2.3% of unknown sex and age. The skin tone distribution on the Fitzpatrick scale consists of 23.4% tone I, 38.3% tone II, 14.9% tone III, 9.8 tone IV, 7.4% tone V, and 6.12% tone VI. The age groups distribution is 9.3% children, 25.8% young adults, 41.1% adults and middle-aged individuals, and 21.5% seniors. FAIRSET annotations do not match one-to-one with Alexa WiderFace labels, but the variation extent is not a major issue since skin tone estimation is subjective and can vary slightly with lighting, background, and camera. The FAIRSET annotations underwent multiple validation steps to ensure data quality, with 86% of skin tones within one degree of offset. The Alexa WiderFace dataset serves

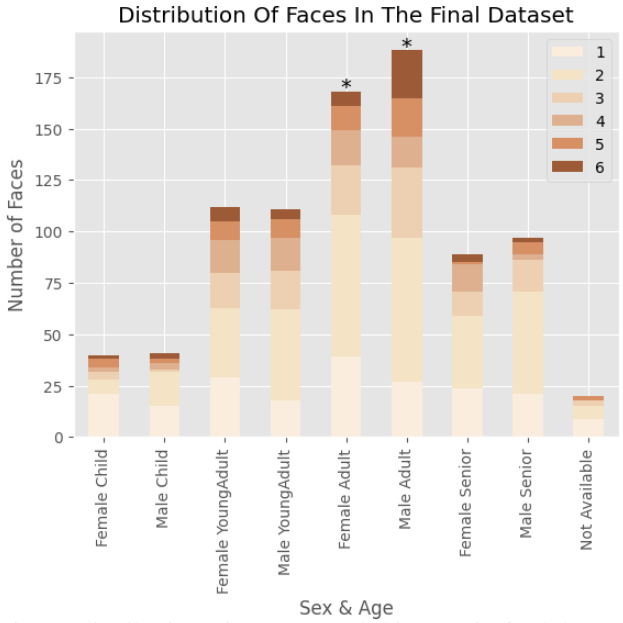


Fig. 3: Distribution of sex, age and skin tone in final dataset
*Recall that *Adult* and *Middle-aged* groups were combined

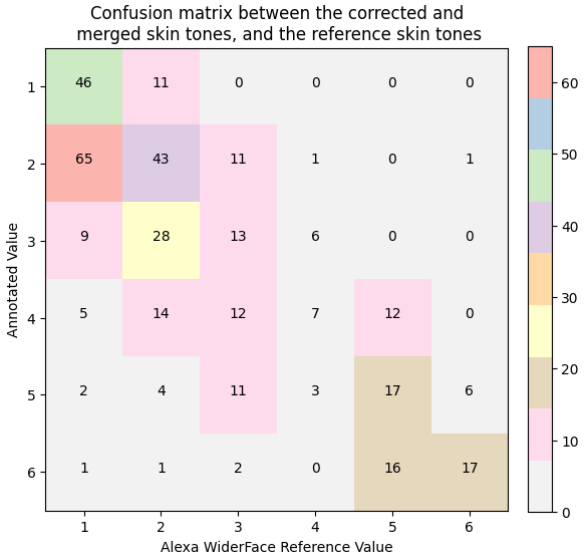


Fig. 4: Confusion matrix between the final skin tone annotations and the Alexa WiderFace reference labels

more as a reference than a ground truth, and both source labels and FAIRSET annotations may contain some errors. It should also be pointed out that the given *sex* labels are subjective and may not represent the gender identity or sex at birth of the individuals. The NME metric, in turn, gives a good picture of the relative inter-annotator precision between keypoints, and it can be translated to a real-world error estimate. Using an average North American white young adult male with an interocular distance of 3.29 cm [12], facing the camera, and enclosed by a bounding box of 400x550 px, an NME magnitude of 0.1 would give a real-world error of 0.33 cm, while a magnitude of 0.5 would increase this error to 1.65 cm. Finally, adding videos in a subsequent iteration of the dataset might refine FLD tools evaluation, as some, such as MediaPipe, apply smoothing and frame-dropping. Despite the stated limitations, FAIRSET offers the first face landmarks database that includes demographics information, enabling the evaluation of biases in existing face landmark detection models. FAIRSET is available for scientific and non-commercial use at the Research Center on Aging GitHub's page: https://github.com/CDRV/FAIRSET.

Acknowledgment

This manuscript was reviewed for language and fluidity with the assistance of Microsoft Copilot. This study is supported by the Canada Research Chair in Biomechanical Movement Signature to Better Understand and Reduce Disability. N.Z. is supported by NSERC.

References

[1] Y. Qiu and Y. Wan, "Facial Expression Recognition Based on Landmarks," in *2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Dec. 2019, pp. 1356–1360.

[2] F. Haugg, M. Elgendi, and C. Menon, "GRGB rPPG: An Efficient Low-Complexity Remote Photoplethysmography-Based Algorithm for Heart Rate," *Bioengineering*, vol. 10, no. 2, p. 243, Feb. 2023.

[3] A. Ablavatski, I. Grishchenko, Y. Kartynnik, K. Raveendran, and M. Grundmann, "Attention Mesh: High Fidelity Face Mesh Prediction in Real-time," 2020.

[4] Z. Wang, X. Zhu, T. Zhang, B. Wang, and Z. Lei, "3D Face Reconstruction with the Geometric Guidance of Facial Part Segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 1672–1682.

[5] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "OpenFace 2.0: Facial Behavior Analysis," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, May 2018, pp. 59–66.

[6] A. Khalil, S. G. Ahmed, A. M. Khattak, and N. Al-Qirim, "Investigating Bias in Facial Analysis Systems: A Systematic," *IEEE Access*, vol. 8, pp. 130 751–130 761, 2020.

[7] H. F. Menezes, A. S. C. Ferreira, E. T. Pereira, and H. M. Gomes, "Bias and Fairness in Face," in *2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, Oct. 2021, pp. 247–254.

[8] J. Buolamwini and T. Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender," in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, ser. Proceedings of Machine Learning Research, S. A. Friedler and C. Wilson, Eds., vol. 81. PMLR, 23–24 Feb 2018, pp. 77–91.

[9] K. Zheng, K. Ci, H. Li, L. Shao, G. Sun, J. Liu, and J. Cui, "Heart rate prediction from facial video with masks using eye location and corrected by convolutional neural," vol. 75, p. 103609.

[10] Y. Yang, A. Gupta, J. Feng, P. Singhal, V. Yadav, Y. Wu, P. Natarajan, V. Hedau, and J. Joo, "Enhancing Fairness in Face Detection in Computer Vision Systems by Demographic Bias," in *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Jul. 2022, pp. 813–822.

[11] T. Martyniuk, O. Kupyn, Y. Kurlyak, I. Krashenyi, J. Matas, and V. Sharmanska, "DAD-3DHeads: A Large-scale Dense, Accurate and Diverse Dataset for 3D Head Alignment from a Single," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2022, pp. 20 910–20 920.

[12] L. G. Farkas, M. J. Katic, and C. R. Forrest, "International Anthropometric Study of Facial Morphology in Various Ethnic," *Journal of Craniofacial Surgery*, vol. 16, no. 4, pp. 615–646.

The AI Evolution in Forensic Handwriting Examination: The Road toward Implementation

C. Elisa van den Heuvel
Forensic Research, Faculty of Technology, Amsterdam University of Applied Sciences, Amsterdam, the Netherlands
c.e.van.den.heuvel@hva.nl

Samiah Ibrahim
ForensInk, Ottawa, Canada
samiah@forensink.ca

***Abstract*—**This paper focusses on the intersection of computer science and artificial intelligence (AI) and forensic handwriting examination, and highlights both the remarkable progress already made and the challenges that lie ahead. The journey began with initial collaborations between forensic handwriting examiners (FHEs) and computer scientists (CSs). FHEs voiced their desire for automated systems that could verify subjective interpretations, quantify the strength of evidence, and provide clear, explainable outputs. This sparked a series of competitions at international conferences, challenging AI systems to detect forgeries and analyze handwriting across various languages. From its early stages at the turn of the century to the sophisticated systems of today, one would assume AI is rapidly becoming an indispensable tool in the FHE's arsenal. Where do we stand today? In this presentation, we advocate to start a renewed effort on bringing the fields of FHEs and CSs together to work towards implementation.

Keywords—Forensic handwriting examination, automated signature verification systems, AI, competitions.

I. Pioneering systems: FISH and CEDAR-FOX

The integration of technology into forensic handwriting examination began in the 1980s, marked by collaborations between forensic handwriting examiners (FHEs) and computer scientists. Pioneering systems like FISH (Forensic Information System for Handwriting) and CEDAR-FOX emerged successfully bridging traditional methods with automated analysis. CEDAR-FOX, developed at the University at Buffalo's Center of Excellence for Document Analysis and Recognition [1], assists FHEs by offering interactive processing, writer verification, and searching functionalities, utilizing macro, micro, and style features to determine the likelihood of common authorship. FISH-WANDA (Workbench for Analysis of Numerical Data), an evolution of the FISH system incorporating the WANDA Measurement tool (WAM) [2], focused on improved user interface and portability while maintaining excellent writer identification. Two decades ago, these systems showed promise in streamlining analysis; CEDAR-FOX offered time-saving benefits without replacing human expertise, while FISH-WANDA enhanced feature measurements and user experience.

II. The need for automation: Challenges of traditional forensic handwriting examination

These systems were and continue to be highly promising to help FHEs to overcome the many challenges their traditional working methodology has been facing such as subjectivity, standardization, and practical limitations. With respect to subjectivity, different experts may interpret the same handwriting features in varying ways, leading to inconsistent conclusions. The evaluation process heavily relies on the FHEs knowledge and judgment, which can introduce personal biases and differing opinions. Furthermore, the absence of regulation of existing standards across regions and jurisdictions complicates the comparison of results and the establishment of a cohesive framework for handwriting analysis. This could pose problems in legal proceedings where consistency and reliability are paramount.

Practical limitations further compound the issues in traditional handwriting examination. Manual examination of handwriting samples is a time-consuming process, particularly in complex cases involving numerous documents and multiple writers. This inefficiency can lead to backlogs in forensic laboratories and delays in criminal investigations. Additionally, the potential for examiner bias exists, where preconceived notions or external pressures may influence the analysis. In cases involving disguised writing or limited sample availability, results may be inconclusive, hampering the progress of investigations.

The integration of automated verification tools offers promising solutions to address these challenges [3]. By leveraging machine learning, these tools can enhance objectivity through statistical analysis of writing patterns and standardization through applicating uniform analysis criteria. This approach reduces the impact of subjective biases and improves consistency across examinations. Automated systems can process large volumes of samples quickly, significantly improving efficiency and reducing manual inspection time. Advanced feature extraction and pattern recognition capabilities allow for analyzing microscopic details in writing characteristics, such as stroke pressure variations, line quality, and character spacing. Furthermore, statistical analysis and comparison models can quantify similarities and differences, assess the significance of matching features, and estimate the likelihood of common authorship [4]. These technological advancements have the potential to revolutionize forensic handwriting examination, making it more reliable, efficient, and scientifically robust.

III. Benchmarking progress: Competitions and Collaboration (2009-2015)

In recent years, tools like FlashID [5] have been developed to process large volumes of handwriting, offering automated analysis and comparison capabilities. However, when it comes to signature verification, few tools have been designed for practical application in forensic laboratories. In the years 2009 through 2015, collaboration between FHEs and CSs, has intensified at conferences such as the International Conference on Frontiers in Handwriting Recognition (ICFHR) and the International Conference on Document Analysis and Recognition (ICDAR). At these conferences, FHEs provided tutorials to computer scientists about their needs and realistic casework situations [6; 7]. A series of competitions were organized between 2009 and 2015 [8; 9; 10; 11; 12; 13; 14], presenting realistic casework scenarios and questions for computer scientists to test their systems. The ICDAR 2009 Signature Verification Competition (SigComp09) demonstrated promising results, with automated systems achieving Equal Error Rates (EER) as low as 2.85% for online data and 9.15% for offline data [8]. Comparatively, on the same dataset, FHEs gave incorrect conclusions in 3.13% of the verifications when not allowing for inconclusive answers. The 4NSigComp10 and 4NSigCom12 competitions involved both forged and disguised signatures in the dataset [10; 12]. Most systems struggled with disguised signatures. When disguised signatures were excluded, the systems performance improved to 92% accuracy. SigComp11 encouraged the computer science community to enable their systems to compute likelihood ratios instead of EERs [11]. ICDAR competition results revealed that different systems excel at different tasks.

A collaborative study by researchers from both communities [15] provided a man-machine comparison of the results of the competitions. It is shown that the performance of automated systems was highly comparable with that of FHEs in proficiency testing. It was concluded that different automated systems, just like humans, were better on different data. Overall, the results of the competitions were very encouraging [16]. The authors expressed their hope that automated systems would become even better with time and with access to more forensically relevant data especially involving disguised behaviors. The intensive collaboration between the two fields, enabling to compare FHEs and automated systems responses, revealed that automated systems can assist and objectify the FHEs comparisons and conclusions.

Venue	Competition name	What was tested	#systems	Datasets tested
ICDAR 2009	SigComp09	Signatures: skilled forgeries	18 systems	WANDA (NISDCC) signatures & Dutch NFI signatures
ICFHR 2010	4NSigComp10	Signatures: skilled forgeries & disguised	6 systems	Australian La Trobe signatures
ICDAR 2011	SigComp11	Signatures: skilled forgeries	13 systems	Chinese and Dutch NFI signatures
ICFHR 2012	4NSigComp12	Signatures: skilled forgeries & disguised	5 systems	Australian La Trobe signatures
ICDAR 2013	SigWiComp13	Signatures & handwriting	23 systems	Japanese and Dutch NFI data
ICDAR 2015	SigWiComp15	Signatures & handwriting	40 systems	German, Bengali, Italian data

In conclusion, these initiatives have laid a foundation for the potential practical application of such technologies in forensic handwriting casework. The logical progression would be to implement these systems to assist and complement FHEs in routine casework.

IV. Barriers to adoption of automated verification systems

Retrospective analysis from the early developments of FISH and CEDAR-FOX to the present day in 2025 reveals a notable discrepancy. Despite the persistent need for technological advancements in forensic handwriting examination, automated handwriting verification systems, including CEDAR-FOX, have not achieved widespread adoption among FHEs. Nor has there been reporting of the actual use of the automated systems developed and benchmarked at the competitions that were mentioned in this paper. This gap between available technology and its practical implementation in forensic practice warrants closer examination. The limited integration of these advanced systems into forensic workflows, despite their potential to enhance efficiency and objectivity, suggests the presence of complex barriers.

FHEs face an array of challenges when attempting to incorporate automated systems into their workflows, encompassing technological, methodological, and institutional factors. These barriers significantly impede the broad acceptance and utilization of automated systems in routine forensic casework, despite their potential to enhance efficiency and objectivity.

A. Need for a Synergetic Approach

Balancing the integration of human expertise with AI presents a complex challenge. While AI systems show promise in certain aspects of handwriting examination, they may struggle with nuanced or contextual factors that FHEs can readily identify. Conversely, FHEs may be subject to biases or inconsistencies that AI could help mitigate. However, there is also the

risk of bias in training data, which could lead to unfair or inaccurate results, potentially compromising the integrity of investigations. Developing a synergistic approach that leverages the strengths of FHEs and AI systems, while addressing their respective limitations, remains a key challenge in the field of forensic handwriting examination.

B. Integration Issues

Integration of automated systems into traditional forensic workflows remains a significant challenge. Gaps exist between established forensic methodology and newer automated approaches, complicating their incorporation into existing processes. Despite promising results from competitions, none of the automated systems benchmarked in the competitions have been developed into off-the-shelf software that can be readily downloaded and used in forensic crime labs or other forensic document examination practices. This lack of readily available, user-friendly tools creates a barrier to widespread adoption and integration of automated systems in forensic handwriting examination.

C. Reliability concerns in Real-World Scenarios

Doubts persist about the accuracy and dependability of automated systems in real forensic scenarios, particularly for complex or unusual cases. While these systems have shown promise in controlled environments and competitions, their performance in diverse, real-world situations remains a concern. FHEs are particularly cautious about relying on automated systems for cases involving disguised handwriting, limited sample sizes, or other challenging conditions that are common in actual forensic work. Current systems may lack the adaptability required to handle variations in handwriting due to disguise, stress or non-standard writing conditions, limiting their reliability in real-world scenarios. The need for thorough validation and testing of these systems in authentic forensic contexts continues to be a critical challenge.

D. Data Scarcity and Quality

Forensic handwriting examination often faces a critical shortage of high-quality, diverse data. This scarcity stems from privacy concerns, legal restrictions, and the sensitive nature of forensic evidence. Obtaining a wide range of authentic samples, especially those representing various writing conditions and disguised handwriting, remains a significant hurdle. The limited availability of such data hampers the development and testing of robust automated systems.

E. Legal Admissibility Challenges

Courts typically approach AI-generated evidence with caution, often raising concerns about its reliability, transparency, and compliance with existing evidentiary standards. One significant hurdle is the skepticism surrounding the "black box" nature of many AI systems, which makes it difficult, not only for FHEs but also for judges and juries, to assess how the evidence was generated or whether it is free from bias or error. Courts require expert testimony to validate the authenticity and accuracy of AI-generated evidence. Additionally, legal frameworks demand that such evidence demonstrate general acceptance within the relevant scientific community, which is not always achievable for novel AI techniques. Consequently, the admissibility of AI-generated handwriting examination in court proceedings remains a contentious issue, requiring extensive validation, peer review, and establishment of error rates before widespread acceptance in the legal system can be achieved.

Overcoming these challenges necessitates several next steps to refine verification algorithms, develop more comprehensive and diverse datasets, and create systems that seamlessly integrate AI capabilities with forensic handwriting examination. It is crucial to view AI as a complementary tool that enhances, rather than replaces, the skills of trained FHEs, ensuring a balanced and robust approach to handwriting examination in forensic investigations.

V. The path forward: bridging the gap

The future of forensic handwriting examination lies in successfully combining traditional expertise with technological advancements, enhancing human capabilities through objective, quantifiable analysis tools while maintaining rigorous forensic standards. The integration of AI into FHEs casework has been notably limited to date, and practical application in the field remains scarce. However, it is imperative for FHEs to acknowledge the potential impact of AI and proactively guide relevant research, ensuring that AI developments align with the practical needs and ethical considerations of forensic handwriting examination.

The future lies in continued collaboration between FHEs and CSs, with an emphasis on practical implementation while acknowledging current limitations. The authors call for participation and collaboration in ICDAR 2025/2026 to underscore the ongoing nature of this work and the importance of continued international collaboration in advancing the field. Elaborative approaches on the topic of how to effectively bridge the gap are not new [3; 4; 15] and are awaiting the establishment of a common terminology so that implementation can begin.

References

[1] S. N. Srihari, H. Srinivasan, and K. Desai, "Questioned Document Examination using CEDAR-FOX," J. of Forensic Document Examination, 18, 2007, pp. 1-20.

[2] M. van Erp, L. Vuurpijl, K. Franke, and L. Schomaker, "The Wanda Measurement Tool for Forensic Document Examination," J. Forensic Document Examination, 2018, vol. 28, pp. 5-14.

[3] M. Diaz, M. A. Ferrer, D. Impedovo, M. I. Malik, G. Pirlo, and R. Plamondon, "A Perspective Analysis of Handwritten Signature Technology," ACM Comput. Surv., 2019, vol. 51, pp. 1-39.

[4] M. Díaz, M. A. Ferrer-Ballester, and G. Vessio, "Explainable offline automatic signature verifier to support forensic handwriting examiners," Neural Comput. Appl., 2023, vol. 36, pp. 2411-2427.

[5] J. J. Miller, R. B. Patterson, D. T. Gantz, C. P. Saunders, M. A. Walch, and J. Buscaglia, "A set of handwriting features for use in automated writer identification," J. of Forensic Sciences, 70(1), 2025, pp. 1-15.

[6] K. Franke and C. E. Van den Heuvel, "Computational forensics: techniques, applications and challenges in the computed-based analysis of questioned documents," in Proc. 9th Int. Conf. Document Anal. Recognit., Curitiba, Brazil, 2007.

[7] S. Ibrahim and C. E. Van den Heuvel, "Forensic Document Examiners approach to handwriting comparisons/identification," in Proc. 10th Int. Conf. Document Anal. Recognit., Barcelona, Spain, 2009.

[8] V. L. Blankers, C. E. Van den Heuvel, K. Y. Franke, and L. G. Vuurpijl, "The ICDAR 2009 signature verification competition," in Proc. 10th Int. Conf. Document Anal. Recognit., Barcelona, Spain, 2009, pp. 1403-1407.

[9] M. Liwicki, C. E. Van den Heuvel, B. Found, and M. I. Malik, "Forensic Signature Verification Competition, 4NSigComp2010 - Detection of Simulated and Disguised Signatures," in Proc. 12th Int. Conf. Frontiers Handwriting Recognit., Kolkata, India, 2010, pp. 715-720.

[10] M. Blumenstein, M. A. Ferrer and J. F. Vargas, "The 4NSigComp2010 Off-line Signature Verification Competition: Scenario 2," 12th International Conference on Frontiers in Handwriting Recognition, Kolkata, India, 2010, pp. 721-726.

[11] M. Liwicki, M. I. Malik, C.E. van den Heuvel, X. Chen, C. Berger, R. Stoel, M. Blumenstein, and B. Found, "Signature Verification Competition for Online and Offline Skilled Forgeries (SigComp2011)," in *Proc. 11th Int. Conf. Document Anal. Recognit.*, Beijing, China, 2011, pp. 1480-1484.

[12] M. Liwicki, M. I. Malik, L. C. Alewijnse, C. E. Van den Heuvel, and B. Found, "ICFHR2012 Competition on Automated Forensic Signature Verification (4NSigComp 2012)," in Proc. Int. Conf. Frontiers Handwriting Recognit., Bari, Italy, 2012, pp. 823-828.

[13] M. I. Malik, M. Liwicki, L. Alewijnse, W. Ohyama, M. Blumenstein, and B. Found, "ICDAR 2013 Competitions on Signature Verification and Writer Identification for On- and Offline Skilled Forgeries (SigWiComp 2013)," in Proc. Int. Conf. Document Anal. Recognit., Washington, DC, USA, 2013, pp. 1477-1483.

[14] M. I. Malik, S. Ahmed, A. Marcelli, U. Pal, M. Blumenstein, L. Alewijns, and M. Liwicki, "ICDAR2015 competition on signature verification and writer identification for on- and off-line skilled forgeries (SigWIcomp2015)," in Proc. 13th Int. Conf. Document Anal. Recognit., Tunis, Tunisia, 2015, pp. 1186-1190.

[15] M. I. Malik, M. Liwicki, A. Dengel, and B. Found, "Man vs. Machine: A Comparative Analysis for Forensic Signature Verification," J. of Forensic Document Examination, 24, 2014, pp. 21-35.

[16] M. I. Malik, "Automatic signature verification: Bridging the gap between existing pattern recognition methods and forensic science," Ph.D. dissertation, Dept. Comput. Sci., Technische Universität Kaiserslautern, Kaiserslautern, Germany, 2015.



Session

Graphonomics for Education

Comparative Study of Shape and Velocity Reproduction in Neurodevelopmental Disorders

Pablo Santana-Susilla[1], Antonio Parziale[2], M. Sol Fortea[1], Angelo Marcelli[2], Cristina Carmona-Duarte[1]
[1]Universidad de Las Palmas de Gran Canaria, Spain
[2]DIEM, University of Salerno, Italy
pasantanasu@hotmail.com, anparziale@unisa.it, sol.fortea@ulpgc.es, amarcelli@unisa.it, cristina.carmona@ulpgc.es

***Abstract*— The increasing of Attention Deficit Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD) highlights the need for enhancing screening techniques. This preliminary study aims at exploring automated screening based on online handwriting analysis. Two characteristics are evaluated to assess and compare the motor and cognitive skills of ADHD, ASD, and neurotypical children: the ability to complete square drawings in a repetitive copy task and the similarity, among subjects of the same group, of the velocity profiles recorded when repeated squares are drawn. The experiments involved a group of 235 individuals (27 ASD, 14 ADHD and 194 Control) from the age of 4 years to 12 years. Results show an increase in the number of completed squares in the group as age increases, with differences between children with typical development and those with neurodevelopmental disorders. Regarding speed profiles, individuals with ADHD show significant functional differences compared to the other two groups.**

Keywords*— *neurodevelopmental disorder, ADHD, ASD, handwriting analysis

I. Introduction

Autism Spectrum Disorder (ASD) and Attention Deficit Hyperactivity Disorder (ADHD) are neuropsychiatric conditions that significantly affect the quality of life of those who suffer from them. Although they share some clinical manifestations, they have fundamental differences in communication, sustained attention, and motor coordination. Early and accurate detection of these disorders is crucial to implement effective interventions to improve the prognosis and the patient's quality of life. Recent advances in computer vision technologies and data analysis have made it possible to explore how handwriting characteristics may reflect neuropsychological differences between ASD and ADHD. Writing, as a complex motor activity, provides an ideal terrain for identifying distinctive patterns. Children with ADHD often show variability in writing speed due to impulsivity and inattention [1], while children with ASD exhibit more consistent but rigid strokes, reflecting motor rigidity [2, 3].

Despite advances in the knowledge of these disorders, differential diagnosis remains a challenge for healthcare professionals, especially in cases where symptoms overlap. This scenario highlights the importance of technological tools complementing traditional diagnostic methods, allowing for a more objective and quantitative assessment. In addition, the inclusion of factors such as gender and age in the analysis is crucial, as clinical manifestations can vary significantly between groups. In this context, this paper introduces an application designed to analyze square drawings made on digital tablets. Using computer vision algorithms and statistical analysis techniques, the tool identifies key differences in the speed and trajectory profile of handwriting. The results not only offer valuable support for diagnosis, but also provide new perspectives for the study of motor skills in these disorders.

This article describes the development, the methodology used and the results obtained, highlighting the potential clinical impact of this tool in professional practice. In addition, its applicability in real clinical scenarios and its contribution to the advancement of knowledge in the field of child neurodevelopment are discussed.

II. Method

A. Participants

The dataset utilized in this study consists of structured handwriting samples collected from participants diagnosed with ASD, ADHD, and neurotypical controls (C). The database covers people from the age of 4 years to 12 years. All participants were enrolled in school levels appropriate to their chronological age, minimizing variability in educational exposure. Table I details the number of participants by age, diagnosis, and sex. The total number of participants was 235 (27 ASD, 14 ADHD and 194 Control).

Age (years)	Control		ADHD		ASD	
	Male	Female	Male	Female	Male	Female
4	6	3	0	0	3	3
5	3	1	3	0	2	3
6	9	6	2	0	4	0
7	19	16	0	2	0	1
8	25	27	1	1	2	1
9	7	7	2	0	1	2
10	22	17	0	2	2	1
11	7	5	1	0	1	0
12	8	6	0	0	1	0

TABLE I: Distribution of subjects by age, gender, and experimental group

B. Copy Task

After an analysis of the neuropsychological tests adopted by experts, we instructed the participants to copy a square. In fact, drawing a square is included in the NEPSY-II series of neuropsychological tests to assess visuospatial functions of children [4] and it is also included in the Bruininks-Oseretsky Test of Motor Proficiency [5] to evaluate children's fine motor integration. Moreover, according to [6] motor and sensory skills related to drawing a square appear at the age of 4.5, and therefore we can assume that all subjects involved in our study had acquired it at the time of the data acquisition.

C. Data recording procedure

Handwriting data was collected using an iPad Pro (5th generation or higher) with a screen size of 12.9 inches, which was used to display the task to children and record their handwriting. The pencil was provided to them in a neutral position, allowing them to hold it with the hand they felt most comfortable using. The participants were comfortably seated behind a desk, and the experimenter who administered the test was on the side. They were instructed to copy a square shown on the tablet at least once and as many times as they wanted while the system recorded their writing movements.

Each recorded trace was stored as a file, capturing the (X, Y) coordinates and timestamps at each sampling point. This procedure enabled a precise analysis of the writing dynamics, facilitating subsequent computational processing. The dataset was subsequently preprocessed to remove artifacts such as unintended strokes, hesitations, or incomplete figures, ensuring high-quality input for computational analysis.

D. Square Detection

The evaluation of shape reproduction accuracy was conducted through an automated processing pipeline designed to detect squares in digital drawings.

The procedure began with the extraction and preprocessing of coordinate data, ensuring consistency across samples. The raw data were normalized to a 256×256 pixel grid, allowing for standardized analysis regardless of variations in drawing size and position.

Then, smoothing and filtering techniques were applied before performing contour detection to identify closed geometric structures within the drawings. The sequences of 2D points representing the detected contours were approximated into polygons with the Ramer-Douglas–Peucker algorithm [7], an iterative procedure for the polygonal approximation of plane curves. In particular, we used the implementation of the algorithm *approxPolyDP()* made available in the OpenCV library. The algorithm requires the definition of the approximation accuracy ϵ, which is the maximum distance between the original curve and its approximation. We empirically set ϵ equal to 0.015.

A polygon was classified as a square if it exhibited four vertices, convexity, and a minimum area requirement. The minimum square area was empirically set to 1000 units to minimize misclassification of incomplete or distorted shapes. Drawings that did not meet these criteria were excluded from further analysis.

The proportion of correctly drawn squares was computed by calculating the ratio of detected squares to the total number of processed drawings per participant. This approach enabled an objective assessment of the participants' ability to accurately reproduce geometric shapes.

E. Analysis of the velocity profile

Once the tasks successfully completed were identified, that is, those where the subject drew a square, the corresponding velocity profiles were analyzed. Dynamic Time Warping (DTW) was used to measure the similarity of velocity profiles within each diagnostic group (neurotypical, ASD, or ADHD), without cross-group comparisons.

DTW allows for the alignment of time-series data with variations in duration or speed, enabling the identification of common movement patterns despite temporal differences [8]. The DTW distance serves as a measure of similarity between two time series, where a lower DTW distance indicates greater similarity.While DTW is not commonly used in clinical handwriting analysis, it is a standard tool in biometric applications (e.g., signature verification) for comparing temporal patterns.

By comparing DTW distances within each diagnostic group, we assessed whether movement dynamics remained consistent across multiple attempts for the same individual and among different individuals of the same age, sex, and diagnosis. To compare group means, we applied independent sample t-tests with appropriate adjustments for unequal variances using Welch's correction, ensuring robustness despite differences in sample sizes and variance heterogeneity.

III. Results

A. *Percentage of Correct Shape Reproduction in Repetitive Drawing Tasks*

Figure 1 shows the percentage of completed squares between different ages and by sex, comparing control groups (C) with individuals diagnosed with ASD and ADHD. On average, participants contributed approximately 6 valid square drawings each, after artifact filtering and shape validation. These were used for subsequent statistical and velocity analyses.

The percentage of completed squares increases with age in the control group, with a sharp increase at ages 4 to 6, stabilization at 7 to 10, and a subsequent increase. While ASD and ADHD distributions overlap with the control group, ASD individuals show the greatest difficulty, with 85.7% falling outside the Interquartile Range (IQR), compared to 73.3% in ADHD. Females with ASD consistently score lower, remaining below the control group's lower quartile, whereas males show greater variability, with some reaching neurotypical levels. ADHD performance is more variable in early years, especially in males, but stabilizes later. ASD and ADHD distributions overlap with control group trends but show distinct variations in spread and outliers. ASD and ADHD groups exhibit different variability patterns: ASD shows higher spread, indicating a broader range of cognitive performance. ADHD seems more variable in the early years but stabilizes later.

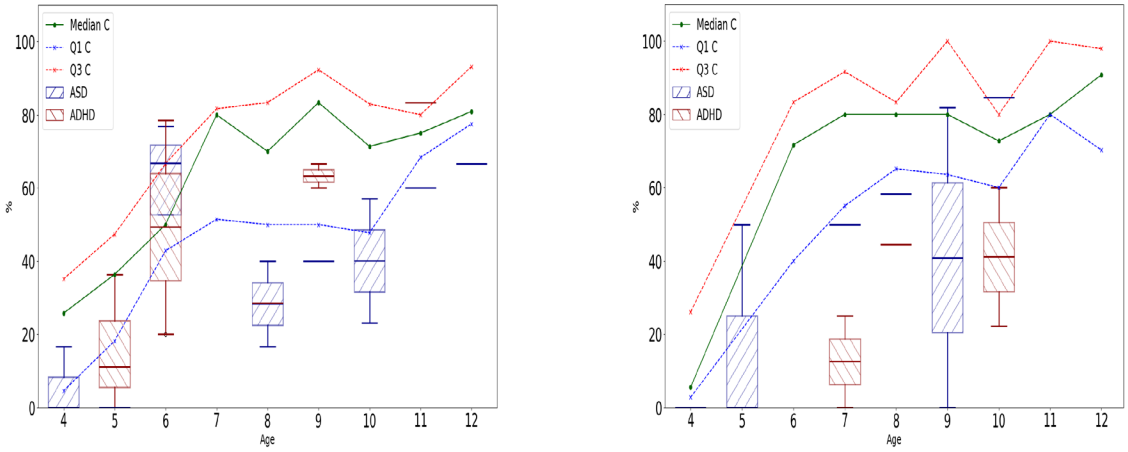


Fig. 1: Comparison Percentage of correct squares in male (a) and female (b) participants.

B. *Differences in the velocity profiles*

Figure 2 shows the distribution of DTW distances in the three groups. The distances were computed between individuals of the same sex and same age.

Because the analysis of the correct shape reproduction in the control group suggested that motor and cognitive learning is more stable in the age bracket of 7–10, only participants aged 7 to 10 were included in the velocity analysis, as the control group's performance stabilised in this age range.

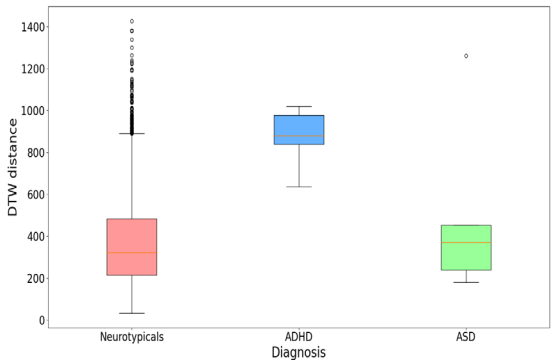


Fig. 2: Distribution of DTW distances between velocity profile in each group.

The comparison of velocity profile similarity within each diagnostic group revealed a significant difference between neurotypical and ADHD children ($t = -8.79$, $p = 0.00031$), indicating that ADHD children exhibit greater variability in their movement patterns compared to neurotypical peers. However, no significant differences were found in velocity profile similarity within the ASD group compared to the neurotypical group ($t = -0.66135$, $p = 0.54452$), nor between the ADHD and ASD groups ($t = 1.83178$, $p = 0.13040$).

These results suggest that while neurotypical and ASD children maintain relatively consistent movement dynamics within their own groups, ADHD children show significantly more variation in their velocity profiles. This aligns with the impulsivity and attention fluctuations characteristic of ADHD, which may lead to less predictable and stable movement patterns.

C. Discussion

The results from shape reproduction accuracy and velocity profile similarity highlight distinct motor control patterns across neurodevelopmental conditions. The percentage of completed squares follows the expected developmental trajectory in the control group, with ASD individuals showing the greatest difficulty, particularly among females, while ADHD individuals exhibit greater variability at early ages but stabilize later.

The ASD group demonstrates persistent difficulties in shape reproduction, with most individuals falling outside the control group's IQR. However, their velocity profiles remain stable when compared to others within their group, suggesting rigid and repetitive motor patterns rather than high variability. In contrast, ADHD individuals, despite performing better in shape reproduction, show significantly greater variability in their velocity profiles when compared to neurotypical children of the same age and sex, aligning with impulsivity and attentional fluctuations characteristic of the disorder.

These findings suggest that motor difficulties in ASD are linked not only to execution errors but also to reduced adaptability in movement patterns, whereas in ADHD, the primary challenge lies in inconsistent motor regulation rather than execution itself. The pronounced difficulties in females with ASD, who remain consistently below the lower quartile of the control group, reinforce the need for sex-specific considerations in diagnostic criteria, as current frameworks may underestimate milder cases in girls. Integrating both motor execution (shape reproduction) and movement pattern similarity (DTW-based velocity analysis) could improve screening tools for neurodevelopmental disorders.

IV. Conclusions and future work

The data obtained in this study provide valuable insights into the differences in fine motor development between children with typical development and those with neurodevelopmental disorders. It is well known that obtaining pediatric samples of ASD and ADHD populations is challenging, but increasing the sample size would be advisable to confirm the observed trends. Due to the limited number of clinical participants, sex-by-diagnosis interactions could not be reliably assessed; this remains a key priority for future studies. Identifying these differences could have a significant clinical impact when making differential diagnoses among various neurodevelopmental disorders. Moreover, designing mobile applications to monitor the fine motor development of children with ASD and ADHD—skills that are crucial during the early years of schooling—could have a major impact on their applicability in real educational settings.

V. Acknowledgement

This work has been supported by the Spanish project PID2021-122687OA-I00 / AEI / 10.13039 / 501100011033 /FEDER, UE.

References

[1] Ruti Cohen, Adi Sasson, and Eynat Gal. Handwriting in children with adhd: The role of graphology. *Journal of Attention Disorders*, 20(3):209–217, 2016.

[2] Casey J Zampella, Leah AL Wang, Margaret Haley, Anne G Hutchinson, and Ashley de Marchena. Motor skill differences in autism spectrum disorder: A clinically focused review. *Current psychiatry reports*, 23(10):64, 2021.

[3] Sol Fortea-Sevilla, Alejandro García-Sosa, Paula Morales-Almeida, and Cristina Carmona-Duarte. Assessment of autism and adhd: A comparative analysis of drawing velocity profiles and the nepsy test. In *21th Conference of the International Graphonomics Society (IGS2023)*, 2023.

[4] M. Korkman, U. Kirk, and S. Kemp. *NEPSY-II Second edition: Clinical and interpretive manual*. Pearson San Antonio, TX, 2007.

[5] Robert H Bruininks and Brett D Bruininks. *BOT-3, Bruininks-Oseretsky test of motor proficiency*. Pearson, 2024.

[6] Thomas A Blondis. Motor disorders and attention-deficit/hyperactivity disorder. *Pediatric Clinics of North America*, 46(5):899–913, 1999.

[7] Urs Ramer. An iterative procedure for the polygonal approximation of plane curves. *Computer graphics and image processing*, 1(3):244–256, 1972.

[8] Stan Salvador and Philip Chan. *FastDTW: Fast Approximation of Dynamic Time Warping*, 2024.

Learning and Transfer of Graphomotor Skills in Typically Developing 7 to 10 Years Old Children and Adults

Laureen Josseron[1], Jérôme Clerc[1], Caroline Jolly[1]
[1]*Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, Grenoble 38000, France*
laureen.josseron@univ-grenoble-alpes.fr

***Abstract*—**Motor transfer is a fundamental aspect of learning. Research shows that children aged 5 to 8 can learn pseudo-letters as effectively as adults. However, it remains uncertain whether their ability to transfer such learning is comparable. In the present study, we investigate the transfer of graphomotor skills in children and in adults. Eighty children aged 7 to 10 and twenty-four adults learned to trace six pseudo-letters from right to left (learning phase). Participants then traced three new pseudo-letters made up of segments learned during the learning phase (first transfer phase). Finally, in the second transfer task, they traced three bigrams and trigrams composed of previously learned pseudo-letters. All tasks were performed on a graphic tablet, which allows to extract kinematic features reflecting the underlying motor processes. We then analyzed the quality of the different handwritten productions as well as three kinematic features: on paper mean speed, fluency (SNVpd feature), and number of abnormal stops. The results show that after the learning phase, children aged 7 to 10 achieved letter quality scores similar to those of adults. In both transfer tasks, all groups displayed a decrease in their speed, fluency, and number of stops. Additionally, 7-year-old children also displayed a decrease in quality scores in the transfer tasks while older children and adults did not. Also, the 8- and 9-year-olds maintained good letter quality in the transfer tasks, but were less fluent and made more stops than adults. This study suggests that after a short learning phase involving fast learning processes, children aged 8 to 10 and adults can form sufficiently abstract representations to enable transfer. In contrast, 7-year-olds have difficulty transferring one's learning, despite slowing down their tracing and making more stops. Finally, children aged 8 and 9 seem to go through a transitional phase in which they maintain a good quality of handwritten productions in the transfer tasks but to the detriment of fluency.

Keywords—graphomotor skills, handwriting, motor learning, transfer

I. Introduction

The acquisition of handwriting is based on a progressive development that begins in early childhood [1]. From the age of two, children start drawing fundamental elements called strokes that are considered as the basic units of writing. These initial graphic productions, although rudimentary, constitute an essential step in acquiring graphomotor skills. Once entering primary school, learning evolves towards letter formation, by combining strokes. The assembly of letters then enables the construction of words, involving more elaborated motor control and increased coordination. This process involves a progressive complexity of motor sequences, highlighting the sequential nature of writing acquisition.

During development, a modification in the mode of motor control is observed. Between the ages of 8 and 9, children progressively switch from an online retroactive control mode, primarily based on sensory feedback, especially visual, to a proactive control mode, which relies on the retrieval of consolidated internal models [1]. This evolution reflects a shift from real-time controlled writing to the progressive automation of handwriting. However, the nature of the motor units represented in these internal models remains debated. Some studies suggest that the fundamental unit represented in these models could be the stroke, while others propose that the motor unit consists of an assembly of smaller units, such as letters. With development, the appearance of "chunks"—groupings of motor units—has been observed, reinforcing the idea of structured sequential learning [2].

The development of writing acquisition has been extensively documented in the literature [1]. However, writing acquisition does not rely solely on motor skills, it also involves linguistic components. Julius & Adi-Japha [3] proposed a paradigm based on an invented letter task to specifically study motor learning in handwriting. Their results indicate that children aged 5 to 8 progress through similar stages to those observed in adults. However, knowledge about the transfer of handwriting acquisition remains limited. Motor transfer, considered as a key stage of motor learning, is defined as the ability to reuse knowledge to improve learning or executing a new movement with similar parameters [4]. It has been suggested that slow learning processes, involved in long-term consolidation, reduce the flexibility of learning and limit the generalization of representations. Conversely, the initial stages of learning, which involve fast learning processes, promote the development of abstract and flexible representations and thus facilitate transfer [5]. Research on graphomotor transfer has primarily focused on adults, with studies on inter-effector transfer [6], and on children, through the study of transfer between different alphabets [7]. In this study, we examined how graphomotor skills develop in children aged 7 to 10 and in adults when learning pseudo-letters. We also investigated whether this learning transfers to new pseudo-letters composed of different sequences of recently acquired strokes, or to bigrams and trigrams formed from these newly learned pseudo-letters.

 ©2024 IEEE

II. Method

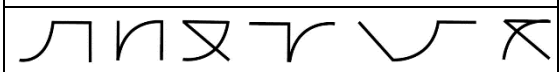
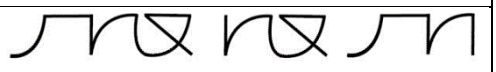
A. Participants

The study included 100 participants: 24 adults (M age: 23.56 years; SD = 7.88; 3 males; 24 right-handed) and 76 children aged 7 to 10 years (from second grade to fifth grade), divided into four groups: 7-year-old children (N = 16; M age: 7.56 years; SD = 0.24; 7 males; 14 right-handed), 8-year-old children (N = 19; M age: 8.50 years; SD = 0.32; 14 males; 17 right-handed), 9-year-old children (N = 20; M age: 9.35 years; SD = 0.28; 10 males, 18 right-handed), and 10-year-old children (N = 21; M age: 10.30 years; SD = 0.25; 12 males; 18 right-handed). All children completed the BHK test [8] to ensure the inclusion of only those without dysgraphia. Four children were excluded because they had dysgraphia. The adults completed the ADC questionnaire [9] to exclude motor and handwriting difficulties.

B. Materials

We designed 12 pseudo-letters, each composed of three segments inspired by Seyll and Content [10], all pseudo-letters are presented in Table 1. These pseudo-letters must be written from right to left, in order to create a novel learning situation and avoid the use of automated movements. The pseudo-letters are presented to participants between two lines, with the starting point marked by a red dot. All participants traced the pseudo-letters on a WACOM© graphic tablet (sampling rate: 200 Hz) with an ink pen that leaves a visible ink trace on the page. We recorded the written tracks using the software developed by the CEA-Leti [11], which then allowed to extract kinematic and dynamic features from the written samples.

TABLE I. Pseudo-letters

Learning Phase	Transfer 1 Phase	Transfer 2 Phase
		

C. Procedure

During the learning phase, participants had to copy the pseudo-letter ten times, starting from a red dot. Then, in a learning test phase, participants had to draw the letter twice without the model. This procedure was repeated for all 6 new pseudo-letters in the learning process. Participants were asked to write as similar to the model as possible. Once the learning phase had been completed, participants were asked to perform two transfer tasks. The order in which the pseudo-letters were presented was counterbalanced.

In the first transfer task, participants were presented with three new pseudo-letters, each composed of three strokes that had been practiced in the pseudo-letters of the learning phase. Each stroke in these new pseudo-letters had been trained in at least two different pseudo-letters during the learning phase. In the first session (Transfer 1, Session 1), participants copied each of the three pseudo-letters from the first transfer task twice, starting from a red dot. In the second session (Transfer 1, Session 2), they copied the same pseudo-letters but were required to do so without lifting the pen, introducing an additional constraint.

In the second transfer task (Transfer 2), participants were presented with three bigrams or trigrams, each consisting of a combination of two or three pseudo-letters already learned in the learning phase. In the first session (Transfer 2, Session 1), participants had to copy each combination twice, starting from the red dot. In the second session (Transfer 2, Session 2), participants had to copy all three combinations starting from the red dot and were instructed not to lift their pen while doing so. The order in which pseudo-letters and transfer 1 and transfer 2 tasks were presented was counterbalanced, but the order of the two sessions remained constant across participants.

D. Measures

The handwritten product was assessed by letter quality on an 11-point scale for each letter written by the participant: Resemblance to the model (0-2 points), one end touching the red point (0-1 point), no interrupted connection (0-1 point), starting from the red point (0-1 point), correct proportion (0-1 point), no retouching or erasure (0-1 point), no exceeding line (0-1 point), angles (0-1 point), respected straight lines (0-1 point), respected curves (0-1 point), and an additional criterion regarding the addition of a new segment in transfer task 2: New segment quality (0-1 point). The quality score of each pseudo-letters was assessed by 4 independent judges. Agreement was moderate to strong (judge 1 and 2, ICC = .71, judge 1 and 3, ICC = .51, judge 1 and 4, ICC = .70). The process was assessed with kinematics measures extracted via the Graph Logger software [11]: the mean on-paper velocity, the writing fluency evaluated through the SNvpd measure [12], and the number of stops longer than 35 mseconds [13].

III. Results

For each of the four variables, a 5 (task: learning test and transfer tasks) x 5 (Age: age groups) repeated measures analysis of variance (rm-ANOVA) was applied to the mean letter quality, mean on paper velocity, SNvpd, and the number of stops. An α level of .05 was adopted. The Bonferroni correction was applied to correct for multiple comparisons. Results are presented in Fig 1.

This study was funded by the company foundation OCIRP, France.

A. Letter quality

The analysis revealed a significant main effect of age, $F(4, 95) = 9.84$, $p < .001$, $\eta 2p = .29$, a significant main effect of task, $F(4, 380) = 7.31$, $p < .001$, $\eta 2p = .07$, and a significant task X group interaction, $F(16, 380) = 2.75$, $p < .001$, $\eta 2p = .10$. Seven-year-olds had lower letter quality than adults in the Transfer 1 Session 1 task, $p < .001$, $\eta 2 = .90$, Transfer 1 Session 2 task, $p < .003$, $\eta 2 = .91$, Transfer 2 Session 1 task, $p < .001$, $\eta 2 = .86$, and Transfer 2 Session 2 task, $p = .001$, $\eta 2 = .89$.

B. Velocity

The analysis revealed a main effect of task, $F(4, 380) = 33.82$, $p < .001$, $\eta 2p = .26$, indicating that all participants show a drop in velocity between the learning test and the Transfer 1 Session 1 task, $p < .001$, $\eta 2 = .93$, the Transfer 1 Session 2 task, $p < .001$, $\eta 2 = .96$, the Transfer 2 Session 1 task, $p < .001$, $\eta 2 = .95$, and the Transfer 2 Session 2 task, $p < .001$, $\eta 2 = .93$. All other effects were not significant.

C. Fluency (SNvpd)

The analysis revealed a main effect of age, $F(4, 95) = 7.97$, $p < .001$, $\eta 2p = .25$, a main effect of task, $F(4, 380) = 417.79$, $p < .001$, $\eta 2p = .81$, indicating that SNvpd was lower in the Learning test than in all transfer tasks and sessions, $ps < .001$, and SNvpd was lower in Transfer 1 tasks and sessions than in Transfer 2 tasks and sessions, $ps < .001$. Also, the interaction between task and group was significant, $F(16, 380) = 4.56$, $p < .001$, $\eta 2p = .16$. Nine-years-ols had a higher SNvpd than adults in the Transfer 1 Session 2 task, $p = .01$, $\eta 2 = .15$, in the Transfer 2 Session 1 task, $p = .001$, $\eta 2 = .52$, and in the Transfer 2 Session 2 task, $p = .001$, $\eta 2 = .14$. Also, 8-years-olds had a higher SNvpd than adults in the Transfer 2 Session 1, $p = .01$, $\eta 2 = .46$.

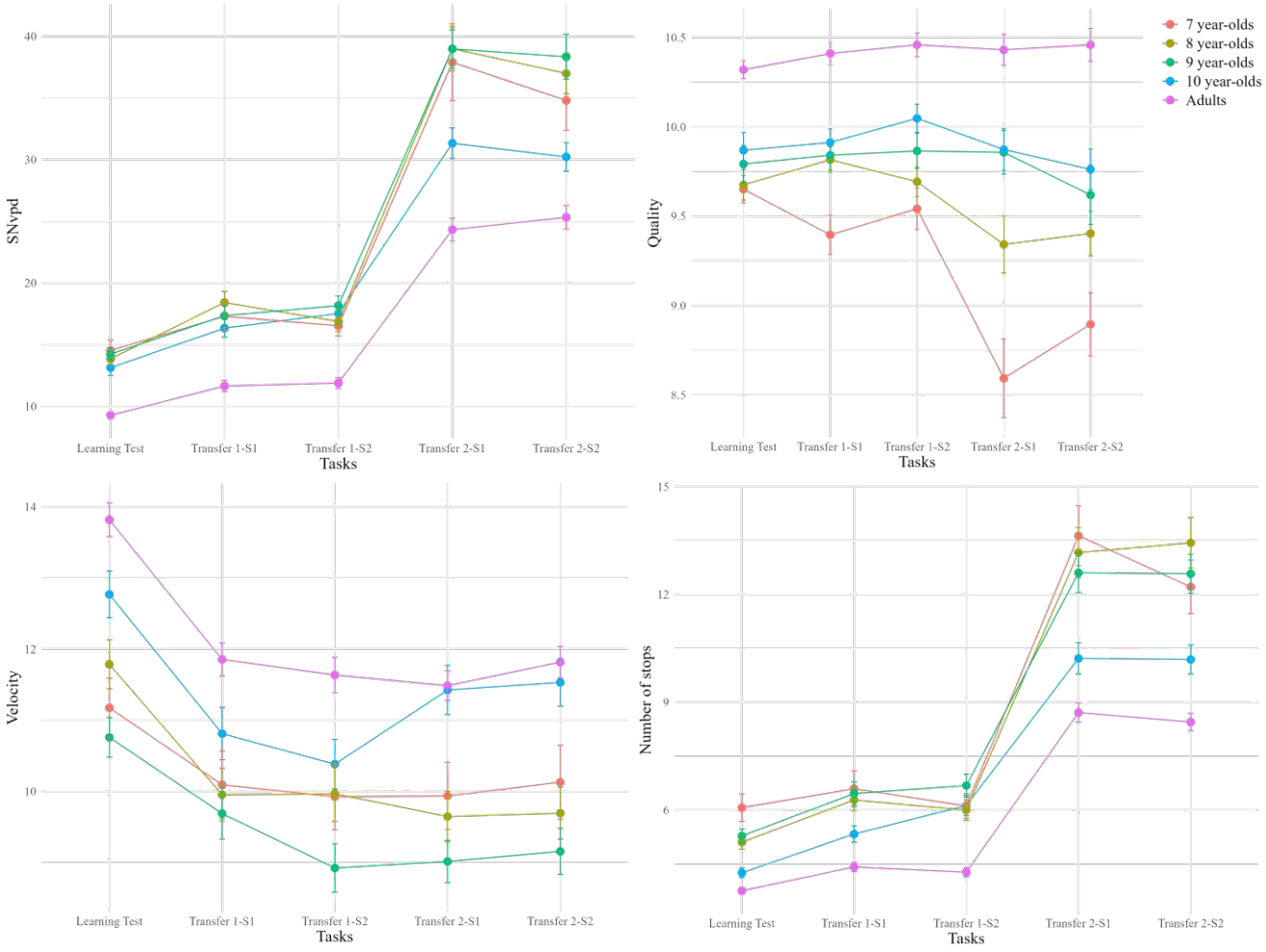


Fig. 1. Graphic representation of Quality, SNvpd, Velocity, and Number of stops for Learning Test and transfer tasks for 7 year-olds, 8 year-olds, 9 year-olds, 10 year-olds and adults. Transfer 1-S1 = Transfer task 1 Session 1 ; Transfer 1-S2 = Transfer task 1 Session ; Transfer 2-S1 = Transfer task 2 Session 1 ; Transfer 2-S2 = Transfer task 2 Session 2.

D. Number of stops

The analysis revealed a main effect of age, $F(4, 95) = 5.80$, $p < .001$, $\eta 2p = .20$. The analysis also revealed a significant main effect of task, $F(4, 380) = 323.72$, $p < .001$, $\eta 2p = .77$, indicating that the number of stops was lower in the Learning test than in all transfer tasks and sessions, $ps < .001$, and the number of stops was lower in Transfer 1 tasks than in Transfer 2 tasks, $ps < .001$. The interaction between task and group was significant, $F(16, 380) = 3.80$, $p < .001$, $\eta 2p = .14$. Mean number of stops was similar between all age groups in Learning Test, however 9-years-ols had a higher mean number of stops than adults in Transfer 2 Session 1 task, $p = .03$, $\eta 2 = 0.14$. Also, 8-years-olds had a higher mean number of stops than adults in Transfer 2 Session 2 task, $p = .02$, $\eta 2 = .39$.

IV. Discussion

All groups show similar performance in the learning test, suggesting that, in a short training context involving fast motor learning processes, children aged 7 to 10 perform at the same level as adults. This observation aligns with existing literature [3]. However, this study aimed to characterize the product and process of handwriting during transfer based on participants' age. Letter quality analyses revealed that when two transfer tasks were introduced, 7-year-old children seemed to encounter more difficulty transferring their learning compared to older children and adults. Interestingly, the decrease in speed during the transfer tasks allowed older children and adults to successfully copy the new pseudo-letters, but this was not the case for 7-year-olds, even though their speed had also decreased between the learning test and the transfer tasks. Similarly, the number of stops, an indicator of movement preparation, did not seem to compensate for the adaptation of knowledge to the new transfer tasks. This was particularly evident in the second transfer task, which involves a combination of pseudo-letters learned during the acquisition phase.

Furthermore, fluency decreased for all participants, despite maintained letter quality in transfer tasks for children aged 8, 9, and 10, as well as adults. Interestingly, the speed of 8- and 9-year-old children was comparable to that of adults; however, 9-year-old children exhibited less fluency than adults in the transfer tasks, and their number of stops was greater in the second transfer task. Similarly, at 8 years old, children showed poorer fluency in the second transfer task (Session 1) and a higher number of stops (Transfer 2 Session 1). Thus, it seems that in the 8-9 age range, which corresponds to the transition period from a feedback-based, to a proactive control mode, children are able to write quickly but at the expense of the fluency and the number of stops. These results indicate that graphomotor development during the transfer of learning is not linear, and children aged 8-9 experience a degradation in writing fluency, even if the letter quality is maintained. This deterioration in fluency appears to be transitory, since 10-year-olds achieve the same performance as adults. Thus, for children aged 8 to 10 and adults, it seems that the early phases of learning, which rely on fast learning processes, allow for the creation of a motor representation sufficiently flexible to be transferred. For 7-year-olds, transfer is more difficult, and children aged 8-9 maintain good writing quality in the transfer task but at the cost of fluency.

The fluency and the number of stops seemed to be less affected by the first transfer task than by the second for all participants. The first transfer task involved a new sequence of known segments, while the second consisted of new combinations of learned sequences. This suggests that at this stage of learning, smaller units, such as segments, are more represented in internal models than sets of units. However, a limitation of this study is that the second transfer task, composed of bigrams and trigrams of known pseudo-letter combinations, may be more difficult to perform because it contains more segments. The length of the combinaisons of pseudo-letters may have been an additional challenge for the younger children, which could have affected the pseudo-letter quality in this task. Future studies controlling for the difficulty of transfer tasks would help confirm these findings.

Additionally, further research could evaluate the transfer abilities of children and adults after extended training to characterize graphomotor transfer across all learning phases. Transfer tasks could also be varied to assess both near and far transfer at all stages of motor learning and new transfer tasks to assess transfer to different strokes sequences and writing styles can confirm these results. Finally, a study [14] has shown that children with developmental coordination disorder, who are known to have difficulties in creating and consolidating internal models, exhibit effective motor learning but weaker transfer performance compared to their peers when writing a new letter. Investigating the ability of such children to transfer their learning during the early stages of learning could improve our understanding of their difficulties and inform targeted interventions [15].

References

[1] Palmis, S., Danna, J., Velay, J. L., & Longcamp, M. (2019). Motor control of handwriting in the developing brain: A review. *Developmental Dysgraphia*, 123-140. https://doi.org/10.1080/02643294.2017.1367654

[2] Diedrichsen, J., & Kornysheva, K. (2015). Motor skill learning between selection and execution. *Trends in Cognitive Sciences*, *19*(4), 227-233. https://doi.org/10.1016/j.tics.2015.02.003

[3] Julius, M. S., & Adi-Japha, E. (2015). Learning of a simple grapho-motor task by young children and adults: similar acquisition but age-dependent retention. *Frontiers in Psychology, 6*, 225. https://doi.org/10.3389/fpsyg.2015.00225

[4] Seidler, R. D., & Noll, D. C. (2008). Neuroanatomical correlates of motor acquisition and motor transfer. *Journal of Neurophysiology, 99*(4), 1836-1845. https://doi.org/10.1152/jn.01187.2007

[5] Clark, D., & Ivry, R. B. (2010). Multiple systems for motor skill learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(4), 461-467. https://doi.org/10.1002/wcs.56

[6] Andree, M. E., & Maitra, K. K. (2002). Intermanual transfer of a new writing occupation in young adults without disability. *Occupational Therapy International*, *9*(1), 41-56. https://doi.org/10.1002/oti.155

[7] Asselborn, T., Johal, W., Tleubayev, B., Zhexenova, Z., Dillenbourg, P., McBride, C., & Sandygulova, A. (2021). The transferability of handwriting skills: from the Cyrillic to the Latin alphabet. *npj Science of Learning*, *6*(1), 6. https://doi.org/10.1038/s41539-021-00084-w

[8] Charles, M., Soppelsa, R., & Albaret, J.-M. (2003). BHK – Echelle d'évaluation rapide de l'écriture chez l'enfant. Paris: Editions et Applications Psychologiques.

[9] Kirby, A., Edwards, L., Sugden, D., & Rosenblum, S. (2010). The development and standardization of the adult developmental co-ordination disorders/dyspraxia checklist (ADC). *Research in developmental disabilities*, *31*(1), 131-139. https://doi.org/10.1016/j.ridd.2009.08.010

[10] Seyll, L., & Content, A. (2022). Letter-Like Shape Recognition in Preschool Children: Does Graphomotor Knowledge Contribute?. *Frontiers in Psychology*, *12*, 726454. https://doi.org/10.3389/fpsyg.2021.726454

[11] Deschamps, L., Devillaine, L., Gaffet, C., Lambert, R., Aloui, S., Boutet, J., ... & Jolly, C. (2021). Development of a pre-diagnosis tool based on machine learning Algorithms on the BHK test to improve the diagnosis of dysgraphia. *Advances in Artificial Intelligence and Machine Learning*, *1*(2), 114-135.

[12] Danna, J., Paz-Villagrán, V., & Velay, J.-L. (2013). Signal-to-Noise velocity peaks difference : A new method for evaluating the handwriting movement fluency in children with dysgraphia. *Research in Developmental Disabilities, 34*(12), 4375‑4384. https://doi.org/10.1016/j.ridd.2013.09.012

[13] Paz-Villagrán, V., Danna, J., & Velay, J. L. (2014). Lifts and stops in proficient and dysgraphic handwriting. *Human movement science*, *33*, 381-394. https://doi.org/10.1016/j.humov.2013.11.005

[14] Adi-Japha, E., & Brestel, G. (2020). Motor skill learning with impaired transfer by children with developmental coordination disorder. *Research in Developmental Disabilities*, *103*, 103-671. https://doi.org/10.1016/j.ridd.2020.103671

[15] Josseron, L., Mombo, W. T., Maggiacomo, M., Jolly, C., & Clerc, J. (2025). Transfer of motor and strategy learning in children with Developmental Coordination Disorder (DCD): A scoping review. *Research in Developmental Disabilities*, *157*, 104908. https://doi.org/10.1016/j.ridd.2024.104908

The Impact of Speed and Scale on Spiral Drawing in Children with Developmental Dysgraphia

Jiri Mekyska*†, Jan Mucha*‡, Katarina Zvoncakova†, Zoltan Galaz*, Marcos Faundez-Zanuy‡
*Department of Telecommunications, Brno University of Technology, Brno, Czech Republic
Email: mekyska@vut.cz
†Department of Research Methodology, Institute of Psychology, Czech Academy of Sciences, Brno, Czech Republic
‡Tecnocampus Universitat Pompeu Fabra, Mataró, Spain

***Abstract*— Developmental dysgraphia (DD) is characterized by impaired handwriting fluency and motor control. While graphomotor deficits in DD are well-documented, the impact of movement scale and speed remains underexplored. This study examines how children with DD adapt to drawing spirals of different sizes under explicit time constraints, aiming to identify key motor impairments. We enrolled 269 children (3rd and 4th grade of primary school) who completed two Archimedean spirals (big and small) as quickly as possible. We then extracted features assessing fluency, velocity, and spiral regularity. Statistical and classification analyses were applied to identify key distinguishing features between children with DD and their typically developing peers. Results indicate that children with DD exhibited significantly higher velocity and more frequent velocity fluctuations in the small spiral. A classification model based on small spiral features achieved a balanced accuracy of 61%, with a sensitivity of 77% and a specificity of 45%. Our findings suggest that increased movement speed in DD does not indicate motor fluency but rather reflects reduced control and automation. The significant role of movement scale in graphomotor execution suggests that fine motor deficits in DD become more pronounced in tasks requiring higher precision. These insights contribute to a better understanding of the motor basis of DD and support the development of decision-support systems for graphomotor difficulties.**

Keywords*— *developmental dysgraphia, graphomotor disabilities, spiral, online handwriting, speed, scale

I. Introduction

Developmental dysgraphia (DD) is a specific learning disorder characterized by impaired handwriting fluency and legibility despite adequate cognitive abilities [1]. Affected children struggle with both letter formation and broader graphomotor tasks, exhibiting, for example, increased variability in speed (generally, disfluent handwriting), longer in-air time periods, unstable pressure, etc. [2]. Research consistently shows that children with DD produce writing and drawing movements that are less automated, requiring greater effort and control than their typically developing peers [3], [4]. These motor inefficiencies not only affect academic performance but also impact children's self-esteem and motivation to engage in writing tasks, often leading to avoidance behaviors and frustration [1], [5].

To disentangle motor deficits from linguistic difficulties, studies have examined non-linguistic drawing tasks such as loops and spirals, which engage fine visuo-motor integration without orthographic demands [6], [7]. Children with DD generally show less fluid execution of these elements, with increased velocity fluctuations and motor hesitations. For example, upper loop drawing studies indicate that children with DD struggle to maintain consistent stroke size and spacing, often producing hyper-controlled, rigid movements [6]. Similarly, Archimedean spiral drawing has been identified as a sensitive tool for assessing motor irregularities, with children with DD displaying greater spatial inconsistency and reduced fluency compared to controls [8]. These findings support the notion that DD is not merely a handwriting issue but a broader motor coordination challenge with implications for other fine motor tasks.

However, key gaps remain in the literature. First, while dysgraphic handwriting deficits are well documented, the impact of movement scale (e.g., large vs. small graphomotor elements) has not been systematically explored. For example, drawing a small Archimedean spiral requires fine motor precision, whereas a large one relies more on gross motor coordination. Understanding whether dysgraphia-related impairments are more pronounced at a specific movement scale can provide insight into the nature of motor deficits. Additionally, differences in muscle control and spatial planning may contribute to variations in performance, suggesting that the role of proprioceptive feedback in dysgraphic motor execution warrants further investigation [9].

Second, few studies have examined how children with DD adjust under explicit speed constraints. When asked to draw as quickly as possible, do they fail to speed up, suggesting motor rigidity, or do they increase speed at the expense of accuracy? Addressing this question could provide important insights into whether DD stems from an intrinsic motor timing issue or a compensatory strategy to maintain speed despite impaired fine motor skills.

Therefore, this study aims to address these knowledge gaps by comparing the performance of children with DD and typically developing controls in a quickly drawn Archimedean spiral, focusing on the interaction between diagnosis and drawing scale. We hypothesize that, in comparison to their typically developing peers, children with DD will show greater performance deficits (lower fluency, increased irregularity, and faster execution) when drawing small spirals, as the smaller spiral requires finer motor control, which is particularly impaired in DD.

II. Materials and Methods

A. Database

We enrolled 269 children (age 9.1±0.5 years for the 3rd grade and 10.2±0.5 years for the 4th grade) from Czech primary schools. Participants were stratified into two groups: intact children (IC; 50 in the 3rd grade, 64 in the 4th grade) and those with developmental dysgraphia (DD; 59 in the 3rd grade, 96 in the 4th grade), as diagnosed by a special education needs teacher who followed a comprehensive Czech standardized battery (including achievement tests for school skills). Regarding gender distribution, 34 boys and 16 girls attended the 3rd grade in the IC group, while 50 boys and 14 girls attended the 4th grade in the IC group. In the DD group, 38 boys and 21 girls attended the 3rd grade, and 75 boys and 21 girls attended the 4th grade. No significant differences in age or gender were observed between the groups in either grade.

During data acquisition, children were instructed to complete two Archimedean spirals (7 cm and 14 cm wide, respectively) on a Wacom Cintiq 16 (DTK1660K0B) display tablet as quickly as possible. Each graphomotor element was printed on a separate sheet of paper. Before beginning the acquisition, the children were given an opportunity to familiarize themselves with the tablet by engaging in free-drawing activities, such as drawing a random picture.

The parents of all children participating in this study provided informed consent, which was approved by the Ethics Committee of the Institute of Psychology at the Czech Academy of Sciences. The study adhered to the Ethical Principles of Psychologists and Code of Conduct as outlined by the American Psychological Association [10] throughout its duration.

B. Feature Extraction

To quantify the regularity of the spirals and assess drawing fluency, we extracted the following commonly used features (although primarily in the context of Parkinson's disease dysgraphia) from the recorded time series [11], [12]: the relative number of interruptions (i.e., the number of pen elevations normalized by duration), the degree of spiral drawing severity, first-order smoothness of the spiral, second-order smoothness of the spiral, the spiral precision index, spiral tightness, variability in spiral width, and the relative number of changes in the velocity profile (normalized by duration).

Additionally, to further examine kinematic abilities and analyze finer details in handwriting kinematics, we employed fractional order derivatives (FDs). Fractional calculus (FC) represents a generalization of classical differential calculus and allows us to analyze signals with temporal memory effects and non-integer orders of differentiation. By utilizing FDs, we aim to capture subtle nuances in handwriting dynamics that may not be apparent in traditional integer-order derivatives. In our previous works, we have explored the use of FDs for analyzing handwriting in neurodegenerative disorders [13] and children with DD [14], [15], demonstrating their potential to capture subtle nuances in motor control and coordination. For this study, we utilized the Caputo FD, defined as [16]:

$$ {}^{C}\mathrm{D}^{\alpha} y(t) = \frac{1}{\Gamma(n-\alpha)} \int_{0}^{t} (t-\tau)^{n-\alpha-1} y^{n}(\tau)\, \mathrm{d}t, \tag{1} $$

where ${}^{C}\mathrm{D}^{\alpha} y(t)$ denotes the Caputo derivatives of order α of the function $y(t)$, Γ is the gamma function, and $n-1 < \alpha \leq n, n \in \mathbf{N}, t > 0$.

FD kinematic features were calculated for α orders 0.1, 0.5, and 0.9. Features were computed from on-surface movement only and for each separate stroke. Subsequently, we merged the stroke features into a single representation by calculating the median. Features were computed separately from the trajectory and vertical and horizontal movements, resulting in 9 FD features per task. Additionally, we computed standard velocity features (first-order derivative) from the trajectory and horizontal and vertical movements for comparison.

C. Statistical Analysis and Machine Learning

Before conducting the statistical analysis, we considered grade and gender as confounding factors and removed their effects using linear regression [17]. To examine the discriminatory power of the features in each task separately, we employed the Mann-Whitney U test. In addition, we calculated Spearman's correlation with the diagnosis (IC/DD) to determine whether high or low values of a feature were significantly associated with DD. To control for multiple comparisons, we applied a false discovery rate (FDR) correction. To evaluate the effects of diagnosis and scale (large vs. small spiral) on performance, we conducted a two-way analysis of variance (ANOVA). The model included two independent variables: diagnosis and task, along with their interaction. For the ANOVA, features that did not follow a normal distribution (as assessed by the Shapiro-Wilk test) were subsequently transformed using the Box-Cox method.

Finally, to assess the potential of both tasks in a decision-support system, we conducted a classification analysis using the XGBoost algorithm within a 10-fold cross-validation framework with five repetitions. A Bayesian optimization strategy was employed to fine-tune the hyperparameters, with the Matthews Correlation Coefficient (MCC) serving as the primary optimization metric. The model's performance was evaluated using balanced accuracy (BACC), sensitivity (SEN), specificity (SPE), and MCC. Additionally, to gain insight into the contribution of particular features, we explored feature importances.

III. Results

Results of the exploratory statistical analysis are given in Table I. The table includes only features that significantly differed between the IC and CC groups, as assessed by the Mann-Whitney U test, in at least one task. All significant features are velocity-based. For the big spiral, we observed a significant difference in the number of changes in the velocity profile and vertical velocity; however, these results did not pass the FDR correction. Regarding the small spiral, we identified significant differences in vertical and horizontal velocity, as well as in the number of changes in the velocity profile – all of which remained significant after FDR correction. In both tasks, children with DD exhibited increased velocity and a higher number of changes in the velocity profile.

TABLE I. Selected Results of the Exploratory Statistical Analysis

Big spiral					
Feature	ρ (IC/DD)	p (IC/DD)	adj p (IC/DD)	p (MWU)	adj p (MWU)
relative number of changes in the velocity profile	0.13	**0.0306**	0.2218	**0.0310**	0.2220
vertical velocity (median, α = 0.9)	0.13	**0.0401**	0.2218	**0.0404**	0.2220
vertical velocity (median, α = 1.0)	0.14	**0.0183**	0.2218	**0.0186**	0.2220
horizontal velocity (median, α = 0.1)	0.05	0.3981	0.5307	0.3975	0.5300
horizontal velocity (median, α = 0.5)	0.10	0.0998	0.2218	0.0999	0.2220
horizontal velocity (median, α = 0.9)	0.05	0.4400	0.5491	0.4394	0.5484
Small spiral					
Feature	ρ (IC/DD)	p (IC/DD)	adj p (IC/DD)	p (MWU)	adj p (MWU)
relative number of changes in the velocity profile	0.21	**0.0006**	**0.0042**	**0.0007**	**0.0047**
vertical velocity (median, α = 0.9)	0.16	**0.0103**	**0.0411**	**0.0105**	**0.0422**
vertical velocity (median, α = 1.0)	0.06	0.2990	0.4663	0.2985	0.4655
horizontal velocity (median, α = 0.1)	0.21	**0.0005**	**0.0042**	**0.0006**	**0.0047**
horizontal velocity (median, α = 0.5)	0.23	**0.0002**	**0.0031**	**0.0002**	**0.0037**
horizontal velocity (median, α = 0.9)	0.18	**0.0026**	**0.0132**	**0.0028**	**0.0140**

* ρ (IC/DD) – Spearman's correlation coefficient with the diagnosis; p (IC/DD) – p-value of Spearman's correlation with the diagnosis; adj p (IC/DD) – adjusted p-value of Spearman's correlation with the diagnosis; p (MWU) – p-value of the Mann-Whitney U test; adj p (MWU) – adjusted p-value of the Mann-Whitney U test

The two-way ANOVA revealed significant main effects of both diagnosis and scale on several drawing features. Children with DD exhibited significantly greater degree of spiral drawing severity ($p = 0.0038$), lower tightness of the spiral ($p = 0.0310$), lower spiral precision ($p = 0.0012$), a higher number of changes in the velocity profile ($p = 0.0038$), and increased horizontal velocity ($p = 0.0473$ for $\alpha = 0.9$). The scale also had a significant impact, with degree of spiral drawing severity ($p = 0.0392$), first-order smoothness of the spiral ($p = 0.0359$), and velocity differing between the big and small spiral tasks. This difference was particularly pronounced in vertical velocity (e.g., $p < 0.0001$ for $\alpha = 0.9$), where lower values were observed in the small spiral. No significant interaction effects between diagnosis and scale were found.

Regarding the classification analysis, the model trained on features extracted from the small spiral demonstrated superior performance, achieving a BACC of 0.61±0.09, with SEN of 0.77±0.08, SPE of 0.45±0.16, and MCC of 0.23±0.19. In contrast, the model based on the large spiral obtained a BACC of 0.53±0.08, with SEN of 0.71±0.10, SPE of 0.35±0.13, and MCC of 0.07±0.17. Feature importance analysis revealed that the most relevant features were related to FD-based velocity metrics, with vertical movement features being more represented in the first model (large spiral), while a combination of both vertical and horizontal velocity-based features was observed in the second model (small spiral). Additionally, the spiral precision index and the relative number of changes in the velocity profile contributed significantly to the classification.

IV. Discussion

Children with DD drew spirals faster and with more frequent velocity changes than controls but showed worse quality (greater distortion). This suggests that speed gains did not improve performance. Prior studies similarly report disfluent, effortful strokes in dysgraphia, where increased speed most likely reflects poor motor automation rather than fluency [2], [6]. This highlights a core graphomotor impairment: faster movement comes at the cost of control.

To further explore this, we performed Spearman's correlation between median vertical velocity (α = 0.9; significant in both tasks) and spiral quality. Among typically developing children, higher speeds (especially on the large spiral) correlated with worse quality: greater spiral drawing severity (ρ = 0.34, p = 0.0013) and reduced tightness (ρ = -0.32, p = 0.0013), reflecting a speed–accuracy trade-off. However, children with DD showed no strong speed–quality correlation on the large spiral, with consistently poor performance. Yet, in the smaller, more demanding spiral, increased speed was associated with greater spiral drawing severity (ρ = 0.25, p = 0.0087), suggesting difficulty in regulating speed for accuracy, particularly in fine movements.

Spiral-based classification showed moderate accuracy, with high sensitivity (77%) but low specificity (45%). While this supports its use as a screening tool, the risk of false positives must be considered. Prior studies also report higher sensitivity than specificity in early DD detection [8], suggesting that graphomotor tests can help identify at-risk children for further assessment [18]. Although DD does not always involve generalized fine motor deficits, our findings support that specific motor control challenges can emerge in precision-based tasks such as spiral drawing.

The small spiral (particularly in relation to vertical movement) was especially challenging for children with DD, reinforcing the idea that finer motor coordination is required for small movement amplitudes. Vertical movement engages more muscles and requires greater control, making accuracy more difficult [15], [19].

FD-based features outperformed traditional kinematics in distinguishing children with DD from controls. Fractional-order derivatives capture "memory" effects in motion and detect subtle irregularities missed by integer-order metrics. FD-based velocity measures were more sensitive to dysgraphic anomalies, improving classification. Similar studies report that FD-derived features enhance dysgraphia detection, offering a superior method for quantifying subtle motor impairments [14], [15].

V. Conclusion

This study investigated the impact of speed and scale on spiral drawing in children with DD. Our findings support the hypothesis that children with DD exhibit greater deficits in fluency and motor control, particularly when drawing small spirals, which require finer motor coordination. Compared to typically developing peers, children with DD produced spirals with more frequent velocity changes and increased drawing speed, but at the cost of reduced precision. These results highlight a fundamental motor impairment in DD, where increased speed does not equate to improved performance but rather reflects reduced control and automation.

Acknowledgment

This study was supported by the project Mediated Society (MEDIS:ON) CZ.02.01.01/00/23_025/0008713 co-financed/supported by the European Union and Czech Republic, projects no. 25-15668S and GN23-06074O supported by the Czech Science Foundation, and by grant PID2023-146644OB-I00 funded by MICIU/AEI/10.13039/501100011033 and FEDER EU.

References

[1] P. J. Chung, D. R. Patel, and I. Nizami, "Disorder of written expression and dysgraphia: definition, diagnosis, and management," *Translational pediatrics*, vol. 9, no. Suppl 1, p. S46, 2020.

[2] J. Mekyska, K. Safarova, T. Urbanek, J. Bednarova, V. Zvoncak, J. M. Havigerova *et al.*, "Graphomotor and handwriting disabilities rating scale (GHDRS): Towards complex and objective assessment," *Australian Journal of Learning Difficulties*, vol. 29, no. 1, pp. 1–34, 2024.

[3] T. Gargot, T. Asselborn, H. Pellerin, I. Zammouri, S. M. Anzalone, L. Casteran, W. Johal, P. Dillenbourg, D. Cohen, and C. Jolly, "Acquisition of handwriting in children with and without dysgraphia: A computational approach," *PloS one*, vol. 15, no. 9, p. e0237575, 2020.

[4] S.-H. Chang and N.-Y. Yu, "Handwriting movement analyses comparing first and second graders with normal or dysgraphic characteristics," *Research in developmental disabilities*, vol. 34, no. 9, pp. 2433–2441, 2013.

[5] M. McCloskey and B. Rapp, "Developmental dysgraphia: An overview and framework for research," *Developmental Dysgraphia*, pp. 1–18, 2019.

[6] C. Lopez and L. Vaivre-Douret, "Concurrent and predictive validity of a cycloid loops copy task to assess handwriting disorders in children," *Children*, vol. 10, no. 2, p. 305, 2023.

[7] I. Bosga-Stork, J. Bosga, J. Ellis, and R. Meulenbroek, "Developing interactions between language and motor skills in the first three years of formal handwriting education," *British Journal of Education, Society & Behavioural Science*, vol. 12, no. 1, pp. 1–13, 2016.

[8] J. Mekyska, Z. Galaz, K. Safarova, V. Zvoncak, J. Mucha, Z. Smekal, A. Ondrackova, T. Urbanek, J. M. Havigerová, J. Bednarova *et al.*, "Computerised assessment of graphomotor difficulties in a cohort of school-aged children," in *2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. IEEE, 2019, pp. 1–6.

[9] J. Danna and J.-L. Velay, "Basic and supplementary sensory feedback in handwriting," *Frontiers in psychology*, vol. 6, p. 169, 2015.

[10] A. P. Association *et al.*, "Ethical principles of psychologists and code of conduct," 2016.

[11] M. San Luciano, C. Wang, R. A. Ortega, Q. Yu, S. Boschung, J. Soto-Valencia, S. B. Bressman, R. B. Lipton, S. Pullman, and R. Saunders-Pullman, "Digitized spiral drawing: A possible biomarker for early Parkinson's disease," *PloS one*, vol. 11, no. 10, p. e0162799, 2016.

[12] G. D. Cascarano, C. Loconsole, A. Brunetti, A. Lattarulo, D. Buongiorno, G. Losavio, E. D. Sciascio, and V. Bevilacqua, "Biometric handwriting analysis to support Parkinson's disease assessment and grading," *BMC medical informatics and decision making*, vol. 19, pp. 1–11, 2019.

[13] J. Mucha, M. Gavenciak, J. Mekyska, M. Faundez-Zanuy, L. Brabenec, and I. Rektorova, "Evaluation of fractional calculus and delta parameters in prodromal diagnosis of dementia with Lewy bodies utilizing online handwriting," in *2024 32nd European Signal Processing Conference (EUSIPCO)*, 2024, pp. 1731–1735.

[14] Z. Galaz, J. Mucha, V. Zvoncak, J. Mekyska, Z. Smekal, K. Safarova, A. Ondrackova, T. Urbanek, J. M. Havigerova, J. Bednarova *et al.*, "Advanced parametrization of graphomotor difficulties in school-aged children," *IEEE Access*, vol. 8, pp. 112 883–112 897, 2020.

[15] M. Gavenciak, J. Mucha, J. Mekyska, Z. Galaz, K. Zvoncakova, and M. Faundez-Zanuy, "Computer-aided diagnosis of graphomotor difficulties utilizing direction-based fractional order derivatives," *Cognitive Computation*, vol. 17, no. 1, p. 13, 2025.

[16] M. Caputo, "Linear Models of Dissipation whose Q is almost Frequency Independent—II," *Geophysical Journal International*, vol. 13, no. 5, pp. 529–539, 1967.

[17] M. T. Todd, L. E. Nystrom, and J. D. Cohen, "Confounds in multivariate pattern analysis: theory and rule representation case study," *Neuroimage*, vol. 77, pp. 157–165, 2013.

[18] L. Deschamps, L. Devillaine, C. Gaffet, R. Lambert, S. Aloui, J. Boutet, V. Brault, E. Labyt, and C. Jolly, "Development of a pre-diagnosis tool based on machine learning algorithms on the BHK test to improve the diagnosis of dysgraphia," *Advances in Artificial Intelligence and Machine Learning*, vol. 1, no. 2, pp. 114–135, 2021.

[19] G. P. Van Galen, "Handwriting: Issues for a psychomotor theory," *Human movement science*, vol. 10, no. 2-3, pp. 165–191, 1991.





Author Index

A

Abergel, Lior 88
Akao, Yoshinori 102
Anquetil, Eric 43, 161
Assila, Najoua 138
Attygalle, Nuwan T. 185
Aubin, Carl-Éric 2

B

Bachir, Vanessa 125
Bara, Florence 149
Barchouch, Islam 43
Battista, Amos 25
Begon, Mickaël 39, 138, 166
Belval, Véronique 68
Bencharef, Rayane 144
Berio, Daniel 180
Bernard, Catherine 68
Blanchet, Pierre 125
Bourbonnais, Daniel 106
Brabant, Léanne 68
Braithwaite, Ben 138
Brière, Simon 106

C

Calinon, Sylvain 5, 180
Cardoso Oliveira, Inês 185
Carmona-Duarte, Cristina 35, 125, 227
Cheriet, Mohamed 144
Clerc, Jérôme 231
Coccaro, Rosanna 209
Cordasco, Gennaro 120
Côté-Piché, Victoria 68

D

D'Alessandro, Tiziana 15, 30, 52
Danna, Jérémy 149
De Gregorio, Giuseppe 47
Dehaes, Mathieu 68
Della Cioppa, Antonio 25
De Nunzio, Alessandro Marco 130
De Pandis, Maria Francesca 130
De Stefano, Claudio 15, 30, 130
Diaz, Moises 35, 57, 176, 190, 205
Djioua, Moussa 63, 73, 133, 194
Dundic, Aleksandra 157
Duprey, Sonia 166

E

Espino, David 35
Esposito, Anna 120

F

Fallatah, Hiba 213
Faundez-Zanuy, Marcos 120, 200, 236
Ferrentino, Enrico 209
Ferrer, Miguel A. 35, 57, 176, 190, 205
Fischer, Andreas 157
Fol Leymarie, Frederic 180
Fontanella, Francesco 15, 30, 130
Fortea, M. Sol 227

G

Gaglione, Maria 130
Gagnon, Marie-Michèle 68
Gagnon, Mélanie 68
Galaz, Zoltan 236
Gardella, Marina 79
Girard, Nathalie 43
Graham, Simon J. 4

H

Hamann, Tim 161
Hamden, Mohammed 144
Harbaum, Tanja 161
Harrath, Baha Edine 144
Hosseini-Kivanani, Nina 57
Houde, Sara 111

I

Ibrahim, Samiah 222
Imbert, Florent 161

J

Jimenez-Marcos, Cristina 35
Jithame, Lisa 73
Jolly, Caroline 231
Joly, Ian-Mathieu 93, 218
Josseron, Laureen 231

K

Kämpf, Peter 161
Kang-Dufour, Mi-Suk 68
Kineret Sharfi 20
Klocek, Adam 153
Kreß, Fabian 161

L

Lafrenaye, Sylvie 106
Le Blanc, Flore 106
Lebel, Karina 93, 111, 125, 218
Leiva, Luis A. 57, 185
Léonard, Guillaume 106
Litvak, Marina 88
Luu, Thuy Mai 68

M

Mailly, Robin 39
Marcelli, Angelo 25, 209, 227
Marthot-Santaniello, Isabelle 47
Martin, Camille 106, 111
Mekyska, Jiri 115, 153, 236
Mercier, Lucas 166
Merling, Maor 88
Miatliuk, Konstantsin 190
Molinara, Mario 130
Mucha, Jan 115, 153, 236

N

Naaim, Alexandre 166
Nardone, Emanuele 15, 30, 52
Njomo Njampou, Rostand Marlone 73
O'Reilly, Christian 133

P

Pace, Cesare Davide 52, 130
Parziale, Antonio 209, 227
Paubel, Pierre-Vincent 149
Pelletier, Justine 39
Pena, Rodrigo Cerqueira Gonzalez 47
Plamondon, Réjean 10, 68, 73, 125, 133, 138, 180, 185
Pustovalova, Olga 15

Q

Quintana, José J. 190

R

Rabaev, Irina 88
Regis, Sébastien 10
Remi, Celine 10
Riesco, Eleonor 93, 111, 218
Rodriguez, Cristian 35, 176
Rosenblum, Sara 20
Rozaire, Julie 166

S

Said, Salem 205
Salobrar-García, Elena 57
Santana-Susilla, Pablo 227
Scius-Bertrand, Anna 157
Scotto di Freca, Alessandra 15, 30
Seguin de Broin, Guillaume 68, 125
Segura, Yoan 149
Sehad, Lina 73
Serdyuk, Alexey 161
Serre, Swan 43
Sharfi, Kineret 20
Simard, Marie-Noëlle 68
Simmons III, Seymour 171
Soullard, Yann 161
Suen, Ching Y. 84, 97, 213

T

Tang, Yufang 97
Tardieu, Julien 149
Tremblay, Renaud 111
Trochimczuk, Roman 190
Turner, Craig 39

V

Vaillant, Jean 10
Van, Dinh Dung 47
van den Heuvel, Elisa 222
Verdugo, Felipe 39

W

Wolniakowski, Adam 190

X

Xu, Yan 97
Xu, Yunqi 84

Z

Zelovic, Nikola 93, 218
Zois, Elias N. 205
Zvoncakova, Katarina 115, 153, 236