Biomedical Signal Processing and Control 112 (2026) 108560

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Integrating robotic kinematics and dynamics with online handwriting features for dysgraphia classification

Francesco Brescia [a], Belén Esther Alemán Santana [b], Moises Diaz [b], Gennaro Vessio [a,*], Miguel Angel Ferrer [b], Giovanna Castellano [a]

[a] Department of Computer Science, University of Bari Aldo Moro, Bari, Italy
[b] Institute for Technological Development and Innovation in Communications, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain

ARTICLE INFO

Keywords:
Dysgraphia classification
Handwriting analysis
Deep learning
Feature fusion
Robotic features

ABSTRACT

Dysgraphia, a learning disorder affecting handwriting fluency and legibility, can significantly hamper children's academic development. Early and accurate classification is essential for timely intervention and tailored educational support. This study presents a novel deep learning framework that integrates robotic kinematic and dynamic features from a robotic arm replicating handwriting with traditional online kinematic and temporal features extracted from digitized writing samples. By integrating these complementary features, we aim to enhance dysgraphia classification by capturing detailed handwriting patterns. We transform the multidimensional time-series data into a structured tabular format and process each feature set independently using TabNet, a deep learning model optimized for tabular data. To maximize classification performance, we employ a Tanh-based score fusion strategy, dynamically balancing the contributions of both models. Evaluations on a publicly available dysgraphia dataset demonstrate state-of-the-art performance, achieving 91.7% accuracy and 95.2% precision on the most comprehensive classification task. These results highlight the effectiveness of robotic motion analysis in improving handwriting-based dysgraphia classification, offering a promising tool for clinical and educational screening.

1. Introduction

Handwriting is a fundamental human skill, crucial in communication, education, and cognitive development. Beyond its functional use in daily life, research has shown that handwriting supports literacy acquisition, memory retention, and neural activation related to motor coordination and cognitive processing [1–4]. However, individuals with learning disorders, such as dysgraphia, face significant challenges in acquiring and maintaining handwriting fluency, which can impact their academic performance and self-esteem. Dysgraphia is a neurological condition characterized by impaired motor control during writing, resulting in difficulties with letter formation, spacing, and writing speed [5–7]. Early and accurate identification of dysgraphia is essential for timely intervention and appropriate educational support, as untreated cases can lead to long-term academic struggles and decreased motivation for written tasks.

Traditional dysgraphia assessments rely on qualitative evaluations by specialists, often using paper-based handwriting samples or standardized handwriting tests. While these methods provide valuable insights, they cannot capture the dynamic aspects of handwriting, such as kinematic and temporal features, which are critical for an accurate diagnosis. Researchers have increasingly turned to digital handwriting analysis to address these limitations, leveraging machine learning and deep learning techniques to extract meaningful features from digitized writing samples (e.g., [8,9]). These computational approaches offer objective, repeatable, and scalable assessments that can enhance dysgraphia classification accuracy and reduce the subjectivity associated with manual evaluation.

However, despite the progress enabled by digitizing tablets and styluses, current systems primarily capture surface-level signals — such as pen trajectory, pressure, velocity, and tilt — providing only a partial view of the complex neuromotor processes involved in handwriting. To overcome this limitation, recent work has explored robotic-assisted handwriting analysis as a complementary approach. This paradigm models handwriting as a dynamic process governed by joint articulation and force distribution [10,11]. By integrating robotic kinematics and dynamics, researchers can better approximate human motor functions, leading to a deeper understanding of handwriting difficulties and motor impairments associated with dysgraphia.

This study presents a novel deep learning-based framework for dysgraphia classification that integrates two complementary feature

* Corresponding author.
E-mail address: gennaro.vessio@uniba.it (G. Vessio).

https://doi.org/10.1016/j.bspc.2025.108560
Received 13 March 2025; Received in revised form 14 July 2025; Accepted 7 August 2025
Available online 19 August 2025

Table 1
Summary of related works for dysgraphia classification. The overview includes database type (public or private), participant details (total number of participants and number of dysgraphic participants in parentheses), data type (online/offline), handwriting tasks, extracted features, classifiers, and the best reported performance results.

Ref.	Data	Part.	Type	Tasks	Features	Classifiers	Results
[8]	Priv.	54 (21)	Online	Writing letters	Static, kinematic, non-linear	LDA, RF	Sens: 96%
[14]	Priv.	99 (49)	Online	Writing letters, words, sentences	Kinematic, dynamic, spatial, temporal	SVM	Acc: 90%
[15]	Priv.	298 (56)	Online	BHK test	Static, kinematic, tilt, pressure	RF	Sens: 96.5%
[16]	Priv.	65 (33)	Online	Copy a short paragraph	Kinematic, dynamic, spatial, temporal	RF, SVM	Acc: 84.7%
[17]	Priv.	72 (36)	Online	Writing letters, words, sentences	Kinematic, dynamic, spatial, temporal	SVM	Sens: 75.5%
[18]	Priv.	91 (48)	Offline	Letter writing	Stroke length, velocity, pressure	MLP	Acc: 84.7%
[19]	Priv.	100 (50)	Offline	Writing letters	Kinematic, dynamic, spatial, temporal	CNN	Acc: 85%
[20]	Priv.	76 (15)	Offline	Drawing figures	Kinematic, dynamic, spatial, temporal	XGBoost	Spec: 90%
[21]	Priv.	200 (100)	Offline	Writing letters, words, sentences	Deep features	RF	Acc: 92.6%
[9]	Public	120 (57)	Online	Writing letters, words, sentences	Kinematic, dynamic, spatial, temporal	AdaBoost	Acc: 79.5%
[13]	Public	120 (57)	Online	Writing letters, words, sentences	Kinematic, dynamic, spatial, temporal	SVM	Acc: 80.8%
[22]	Public	120 (57)	Offline	Writing tasks	Deep features	CNN	Acc: 79.7%
[23]	Public	120 (57)	Offline	Writing words and sentences	Deep feature fusion	SVM	Acc: 97.3%

Priv: Private, Acc: Accuracy, Spec: Specificity, Sens: Sensitivity.

sets: (*i*) traditional kinematic and temporal handwriting features extracted from digitized writing samples, including velocity, acceleration, pen pressure, and stroke dynamics; (*ii*) a new set of biomechanical descriptors derived from a robotic arm that replicates handwriting tasks, capturing joint angles and force torques.

These feature sets are processed using TabNet, a deep learning architecture optimized for tabular data that allows dynamic feature selection [12]. Two separate TabNet models are trained independently on the traditional and robotic features, and their outputs are combined using a Tanh-based score fusion strategy, which dynamically balances the contribution of both models to maximize classification performance.

The contributions of this paper are outlined as follows:

- Unlike previous works that rely solely on kinematic handwriting features, this study integrates robotic kinematics and dynamics during writing.
- The proposed pipeline leverages TabNet for feature processing and a Tanh-based score fusion strategy to enhance classification accuracy.
- The method is tested on a publicly available dysgraphia dataset [9], demonstrating state-of-the-art performance with 91.7% accuracy and 95.2% precision on the most comprehensive classification task. Compared to the strongest reported baseline on this task [13], which achieved 80.8% accuracy and 83.3% precision, our model improves accuracy by +10.9% and precision by +11.9%.

The remainder of this paper is organized as follows. Section 2 reviews current approaches to detecting dysgraphia through handwriting analysis. Section 3 outlines the dataset utilized in the study. Section 4 explains the proposed methodology. Section 5 describes the experimental setup and discusses the results, comparing our proposed approach to state-of-the-art methods. Finally, Section 6 summarizes the key findings and suggests directions for future research.

2. Related work

The automatic classification of dysgraphia using handwriting analysis has gained increasing attention in recent years, leveraging machine learning and deep learning techniques to provide objective and scalable solutions. Traditional clinical assessments rely on expert evaluation of handwriting samples, which can be subjective and time-consuming. To address these challenges, researchers have explored various computational methods, initially focusing on kinematic and spatial feature extraction, followed by machine learning-based classification and, more recently, deep learning approaches leveraging online and offline handwriting data (see Table 1).

Mekyska et al. [8] pioneered machine learning for dysgraphia classification by employing Random Forest classifiers on kinematic features such as velocity, acceleration, and pressure. Their approach demonstrated a sensitivity of 96%, underscoring the effectiveness of feature engineering in capturing handwriting anomalies. Building on this, Rosenblum and Dror [14] integrated kinematic and dynamic features to analyze handwriting, achieving 90% accuracy using Support Vector Machines (SVMs). Their study highlighted the importance of combining multiple feature types, such as on-surface and in-air movements, for robust dysgraphia classification.

In subsequent years, feature extraction techniques were refined. Zvoncak et al. [16] introduced fractional derivatives to quantify in-air movements, showing that these features outperformed traditional kinematic metrics in detecting handwriting difficulties. Meanwhile, Dankovičová et al. [17] emphasized the importance of spatiotemporal features, using Principal Component Analysis to enhance interpretability and achieve competitive classification performance with Random Forest models.

The release of a publicly available handwriting dataset by Drotár et al. [9] marked a turning point, enabling broader experimentation with machine learning models. Their study employed AdaBoost classifiers on kinematic and spatiotemporal features, achieving 79.5% accuracy. Kunhoth et al. [13] further explored this dataset, adopting a feature division strategy to reduce dimensionality and improve computational efficiency. Their AdaBoost model achieved an accuracy of 80.8%, slightly surpassing earlier approaches.

In parallel, researchers began integrating deep learning architectures. Skunda et al. [22] converted online handwriting data into offline images to leverage Convolutional Neural Networks (CNNs). Their approach demonstrated the potential of image-based features, achieving an accuracy of 79.7%. Masood et al. [21] extended this idea by combining CNNs with Long Short-Term Memory (LSTM) networks, capturing both spatial and sequential dependencies in handwriting data. Their hybrid model achieved an accuracy of 92.6%, showing the potential of deep learning in this domain.

Kunhoth et al. [23] achieved state-of-the-art performance by leveraging DenseNet201, a deep CNN pre-trained on ImageNet, and fine-tuning it for specific handwriting tasks. They introduced a novel feature fusion strategy that combined task-specific models, reaching an impressive accuracy of 97.3%. This approach highlighted the effectiveness of ensemble methods and pre-trained architectures in handwriting analysis.

Mobile applications have also emerged as a promising avenue for dysgraphia classification, bringing these methods closer to real-world deployment. Kariyawasam et al. [19] developed a mobile app using CNNs to classify handwriting samples, achieving 88% accuracy. Similarly, Asselborn et al. [15] utilized consumer-grade tablets to collect handwriting data, employing Random Forest classifiers to achieve 96.5% sensitivity. These studies underscore the potential for scalable and accessible solutions in dysgraphia screening.

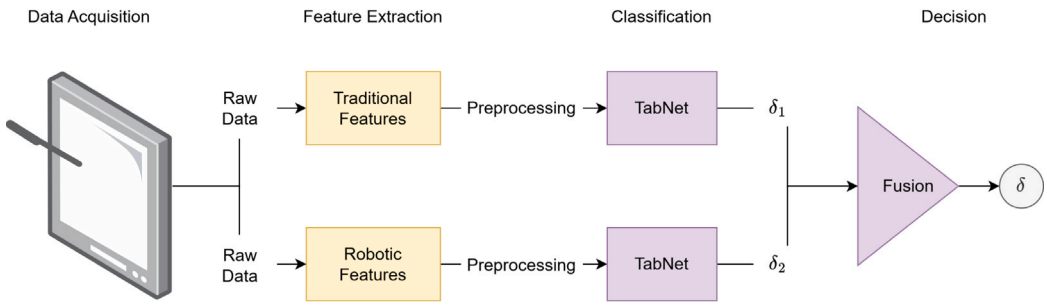

Fig. 1. The proposed framework consists of several key components. Initially, raw data captured by the digitizer is processed to extract robotic and traditional features. These features are then preprocessed and input into two separate TabNet-based models for classification. Finally, a fusion strategy is applied to combine the results, leading to a conclusive decision regarding the classification of dysgraphia in handwriting.

Looking ahead, recent studies in the broader field of clinical AI have explored multimodal Transformer-based architectures for complex diagnostic and prognostic tasks. Nguyen et al. [24] proposed a clinically-inspired multi-agent Transformer framework that mimics the roles of a radiologist and a general practitioner, jointly analyzing imaging and clinical data for disease trajectory forecasting. Similarly, Zhou et al. [25] developed a unified Transformer model that integrates images, structured data, and unstructured clinical narratives for diagnostic support. These architectures demonstrate the capacity of Transformer-based models to capture complex multimodal dependencies and temporal dynamics in medical data. Although such models have not yet been applied to dysgraphia classification, they suggest promising future directions for integrating handwriting trajectories and visual signals within a unified deep learning framework. In the present work, we prioritized model generalizability over architectural complexity, given the dataset size and clinical applicability, thereby laying the groundwork for future multimodal extensions.

Complementary to this direction, recent efforts such as LRAD-ViT [26] have demonstrated that lightweight Transformer models with adaptive token fusion and randomized learning regularization can achieve high diagnostic accuracy while maintaining computational efficiency in medical tasks, including Alzheimer's disease detection. These strategies may inspire future developments in dysgraphia screening models that require a balance between interpretability, resource constraints, and clinical applicability.

A growing body of research suggests that writing is not just a motor skill but involves higher-order control. In this context, robotic emulation of handwriting tasks has emerged as a promising technique to extract biomechanical insights that are not captured by conventional handwriting digitizers. Prior studies have explored the role of robotic features in handwriting analysis, particularly in the context of signature verification [10]. Building on this idea, our study proposes a hybrid approach that combines kinematic and temporal handwriting features with robotic kinematic and dynamic features obtained by simulating handwriting tasks using a robotic arm. Unlike prior work that focuses solely on kinematic analysis, our approach extends the investigation to both kinematic and dynamic robotic features, providing a more comprehensive representation of the handwriting process related to the aspects influencing handwriting impairments.

3. Materials

For this study, the publicly available dataset proposed by Drotár et al. [9] was selected, as it is the only open-access dataset available for dysgraphia classification known to us. The dataset consists of online handwriting data collected from 120 school-age children, 63 of whom were identified as non-dysgraphic, while 57 were diagnosed with dysgraphia. The handwriting samples were acquired using a WACOM Intuos Pro Large digitizing tablet in Slovak orthography. To ensure a natural writing experience, participants wrote on a standard sheet of paper over the tablet while the device recorded handwriting dynamics in real-time.

The raw data include the *timestamp* of each recorded observation, starting from zero for each handwriting session, and the *spatial coordinates* of the pen tip on the writing surface. Additional information captures the *binary status* of the pen, distinguishing between strokes made on the surface and in-air movements. The dataset also includes *pressure* values indicating the force applied during writing, along with angular data such as the *azimuth*, which represents the horizontal orientation of the pen, and the *tilt*, which describes its vertical inclination.

Participants were asked to complete eight handwriting tasks designed to capture different aspects of motor control and cognitive processing. The tasks involved writing the letter *l* and the syllable *le* at both normal and fast speeds, followed by a simple word (*leto*, meaning "summer"), a pseudoword (*lamoken*), a word with diacritics (*hračkárstvo*, meaning "toy shop"), and a complete sentence (*V lete bude teplo a sucho*, meaning "In summer, the weather will be warm and dry"). An additional ninth task, referred to as *all*, was introduced to further analyze handwriting patterns across multiple tasks. This task concatenates all previous handwriting samples into a single, continuous recording, providing a holistic representation of each participant's handwriting style.

4. Methods

This section presents the pipeline developed to achieve the final aim of this study, i.e., detecting dysgraphia in children. The pipeline consists of multiple stages, including extracting traditional and robotic features from raw data, statistical processing, feature selection, and classification using TabNet. A visual representation of the proposed approach is provided in Fig. 1, illustrating how the various components interact within the framework.

4.1. Robotic features

A key contribution of this study is the introduction of robotic features, which augment traditional handwriting data by estimating the underlying biomechanical properties of movement execution. While conventional tablet-based acquisition captures surface-level signals — such as pen trajectory, pressure, and tilt — these measurements primarily reflect the output of the motor system and provide limited insight into the neuromuscular processes responsible for handwriting. In contrast, robotic features provide access to latent motor control dynamics, potentially revealing impairments that are not detectable through standard instrumentation.

To this end, we employed a robotic system that emulates the kinematics and dynamics of a human arm. Using the digitized handwriting trajectories (x, y) from the dataset, the system applies robotic kinematics to compute the corresponding angular positions of the arm joints

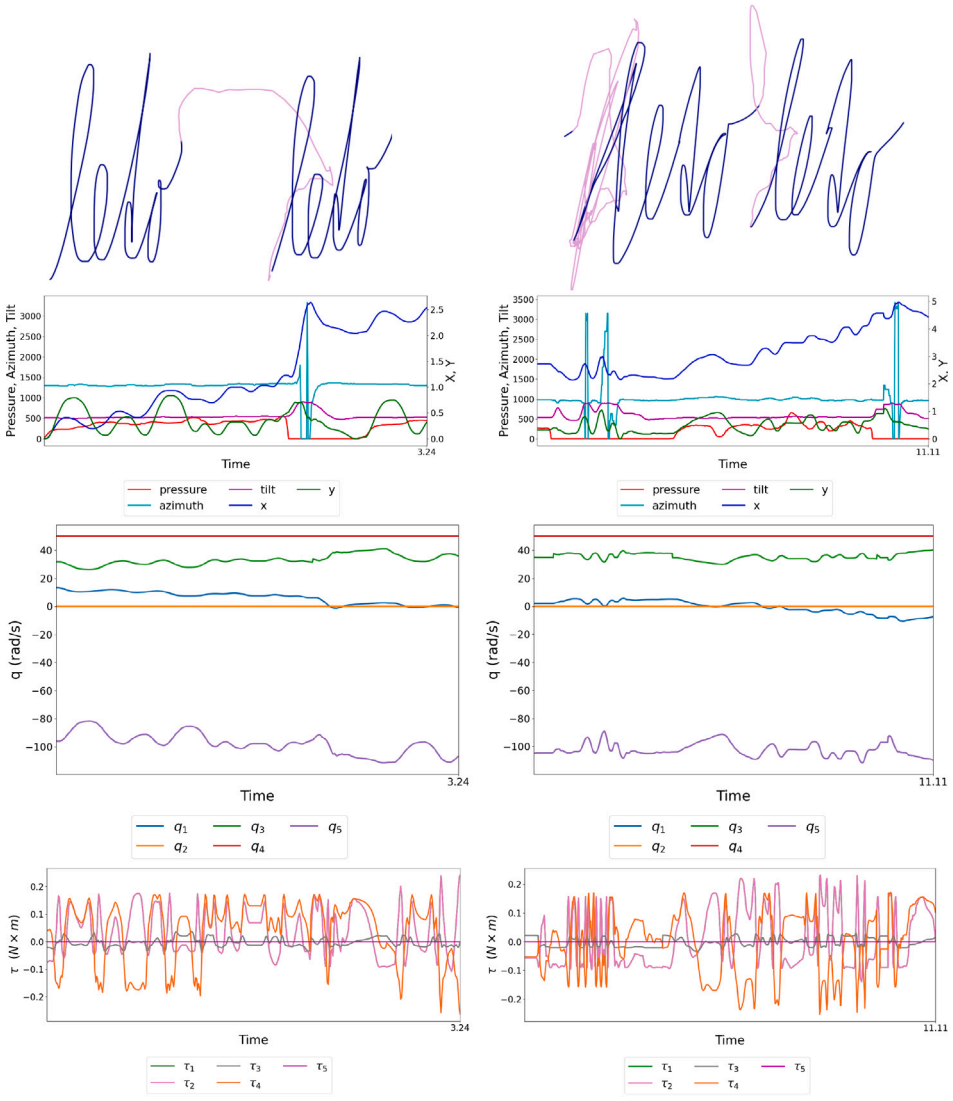

Fig. 2. Example of handwriting samples from a non-dysgraphic (first column) and a dysgraphic (second column) individual. The first row displays the original handwriting, while the second, third, and fourth rows illustrate the corresponding traditional features, and kinematic and dynamic robotic features, respectively.

$(q_1, q_2, q_3, q_4, q_5)$. Subsequently, based on these joint angles, robotic dynamics calculations are performed to estimate the torque forces $(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5)$ required by each joint to perform the writing motion.

This robotic framework enables the extraction of additional motion descriptors that are otherwise unobservable in traditional handwriting analysis. In particular, the robotic-derived features provide insights into the biomechanical effort required for writing and may reveal subtle differences between dysgraphic and non-dysgraphic individuals (Fig. 2). The following sections detail the proposed robotic system's mechanical configuration, kinematic formulation, and dynamic modeling.

4.1.1. Robotic configuration

The robotic model used in this study was designed to closely replicate human arm movements, allowing for the simulation of complex motor tasks such as writing and drawing. The system consists of a robotic arm, which mimics the major articulation in the human arm, pivoted in the elbow. Specifically, our model simulates the forearm, a fingerless hand, and the fingers. To reproduce human joint motion during handwriting, five degrees of freedom are required: one for the elbow, three for the wrist, and one for the knuckle joint. An additional degree of freedom corresponds to the motion of the end-effector, i.e., the pen tip. Accordingly, we define five coordinate frames, S_1 through S_5, each associated with a joint in the kinematic chain. This configuration enables the system to accurately replicate the biomechanical structure of the human upper limb involved in handwriting, allowing for more realistic motion simulation and dynamic feature extraction.

Fig. 3 illustrates the structure of the robotic arm. It highlighting its segments and their anatomical correspondence to the human upper limb, providing a realistic replication of writing kinematics.

The mechanical specifications of the robotic system were derived from anthropometric data to approximate the proportions of an average human arm. Specifically, the forearm and hand constituted approximately 2.3% of total body mass, while the hand alone accounted for 0.7%. The index finger's mass was estimated at 0.01 kg, assuming the total hand weight includes the remaining fingers.

Table 2 summarizes the mechanical properties of each robotic link, including segment lengths and corresponding mass values. These parameters were essential for accurately modeling the system's kinematics and dynamics.

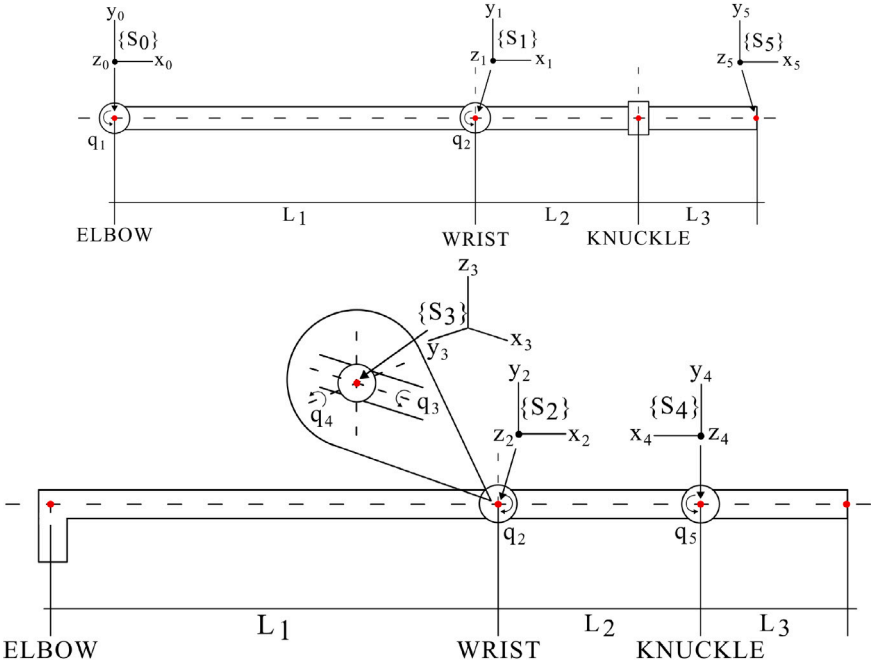

Fig. 3. Illustration of the five-degree-of-freedom robotic arm model from two perspectives: upper view, highlighting one degree of freedom for the elbow joint and one for the wrist joint (top); side view, highlighting three degrees of freedom for the wrist joint and one for the knuckle joint (bottom). The diagram includes the five joint coordinate reference frames, S_1, S_2, S_3, S_4, S_5, with their respective rotations denoted as q_1, q_2, q_3, q_4, q_5. The three links of the robotic arm, labeled as L_1, L_2, and L_3, represent the forearm, hand, and finger, respectively.

Table 2
Mechanical measurements of the links of the proposed robotic model.

Link	Length (cm)	Mass (kg)
L_1 (forearm)	28	1.26
L_2 (fingerless hand)	11	0.48
L_3 (finger)	10	0.01

Table 3
Denavit-Hartenberg parameters of the proposed robotic arm.

Joint	q	d	a	α
1	q_1	0	L_1	0
2	q_2	0	0	$\frac{\pi}{2}$
3	$q_3 + \frac{\pi}{2}$	0	0	$\frac{\pi}{2}$
4	q_4	L_2	0	$-\frac{\pi}{2}$
5	$q_5 - \frac{\pi}{2}$	0	L_3	0

4.1.2. Robotic kinematics

Kinematics describes the relationship between joint positions and the resulting end-effector movements. The primary goal of the kinematic model is to compute the joint angles $(q_1, q_2, q_3, q_4, q_5)$ that enable the robotic arm to trace a given handwriting trajectory in three-dimensional space (x, y, z). Since the digitizers do not directly record the z-coordinate, it is set to zero for on-surface handwriting and 4 mm for in-air handwriting. This inverse mapping is critical for understanding how motor impairments affect writing movement execution.

The Denavit-Hartenberg (D-H) convention [27] was applied to formulate this relationship, defining a standardized approach for representing serial manipulator kinematics. Each joint is characterized by four parameters: rotation angle (q_i), link offset (d_i), link length (a_i), and twist angle (α_i). The kinematic parameters for the robotic arm are outlined in Table 3.

The homogeneous transformation matrices were computed iteratively using the D-H parameters to derive the forward kinematics, enabling the determination of the end-effector position based on joint angles. Conversely, the inverse kinematics problem — determining the joint angles required to achieve a specific end-effector position — was solved using an iterative optimization approach based on the Nelder–Mead simplex algorithm [28]. This method minimizes the Euclidean distance error between the desired and computed end-effector positions.

4.1.3. Robotic dynamics

The dynamic model describes the torques in the joints required to execute the handwriting motion. Estimating these joint torques $(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5)$ provides insight into the biomechanical effort exerted during handwriting, which may differ between dysgraphic and non-dysgraphic individuals.

The Newton-Euler recursive approach was applied to compute the torque values for each joint [29]. This method involves a forward pass to determine linear and angular velocities and a backward pass to compute forces and torques acting on each joint. The estimated torques were further refined using pressure data recorded from the digitizing tablet, following the force-pressure mapping equation proposed by Faundez-Zanuy et al. [30].

4.2. Feature extraction

The dataset provides raw features that offer fundamental insights into handwriting execution, capturing variations in fine motor control that may indicate dysgraphia. However, raw signals alone may not be sufficient for effective classification. Higher-order statistical descriptors are extracted from these features to enhance the discriminative power of the data, transforming the temporal sequences into a structured representation suitable for machine learning models. This step facilitates the identification of patterns associated with handwriting difficulties while reducing noise and dimensionality.

In particular, the extraction process leveraged the open-source library `handwriting features`,[1] which enables the computation of

[1] https://github.com/BDALab/handwriting-features

Table 4
Summary of statistical measures applied to traditional handwriting features and the total number of extracted features. Each feature undergoes statistical summarization to enhance its representational capacity. Features marked with * are computed across three different motion axes (x, y, xy).

Feature	Extracted statistics	Total
velocity_in_air	mean, std, max, min, median, quartile_1, quartile_3	21*
velocity_on_surface	mean, std, max, min, median, quartile_1, quartile_3	21*
acceleration_in_air	mean, std, max, min, median, quartile_1, quartile_3	21*
acceleration_on_surface	mean, std, max, min, median, quartile_1, quartile_3	21*
jerk_in_air	mean, std, max, min, median, quartile_1, quartile_3	21*
jerk_on_surface	mean, std, max, min, median, quartile_1, quartile_3	21*
azimuth_on_surface	mean, std, max, min, median, quartile_1, quartile_3	7
tilt_on_surface	mean, std, max, min, median, quartile_1, quartile_3	7
pressure	mean, std, max, min, median, quartile_1, quartile_3	7
stroke_height_on_surface	mean, std, max, min, median, quartile_1, quartile_3	7
stroke_height_in_air	mean, std, max, min, median, quartile_1, quartile_3	7
stroke_width_on_surface	mean, std, max, min, median, quartile_1, quartile_3	7
stroke_width_in_air	mean, std, max, min, median, quartile_1, quartile_3	7
writing_length_on_surface	–	1
writing_length_in_air	–	1
writing_height_on_surface	–	1
writing_height_in_air	–	1
writing_width_on_surface	–	1
writing_width_in_air	–	1
stroke_duration_on_surface	mean, std, max, min, median, quartile_1, quartile_3	7
stroke_duration_in_air	mean, std, max, min, median, quartile_1, quartile_3	7
writing_duration_on_surface	–	1
writing_duration_in_air	–	1
writing_duration_overall	–	1
number_of_interruptions	–	1
number_of_interruptions_relative	–	1
	Total Features	**200**

a broad spectrum of kinematic, dynamic, temporal, and spatial features derived from handwriting tasks. These features are calculated across different motion axes under conditions similar to those outlined in previous studies [9]. A comprehensive set of statistical descriptors — including mean, standard deviation, maximum, minimum, median and quartiles — is computed to summarize the extracted features effectively. In total, 200 features were obtained using the library (Table 4).

The extraction process for robotic features follows a similar pipeline to traditional handwriting features but incorporates domain-specific computational methods tailored to the novel robotic representations. A preliminary analysis of the robotic dataset revealed that three out of ten features remained constant over time, leading to their removal to eliminate redundant information. To ensure the relevance and robustness of the extracted features, the extraction process was guided by the methodology outlined in [31]. While this study originally focused on signature verification, its preprocessing framework proved well-suited for refining robotic handwriting features.

Building on this approach, twelve time-dependent functions were computed, mirroring those used in [31] to derive additional kinematic, dynamic, geometric, and trigonometric descriptors that enhance the characterization of handwriting movements. These extracted functions were then subjected to the same statistical summarization process applied to traditional handwriting features, ensuring a standardized feature representation and enabling direct comparison between the two modalities. In total, 231 robotic features were derived (Table 5), bringing the final dataset to a comprehensive fusion of traditional handwriting metrics and robotic-based motion analysis.

4.3. Feature normalization and selection

All input variables undergo normalization to guarantee consistency across different scales and units of measurement. This classical standardization process transforms each feature by subtracting its mean and dividing by its standard deviation, preventing features with larger magnitudes from dominating the learning process.

Following normalization, feature selection is performed using a Random Forest classifier [32], a widely adopted ensemble learning method known for its robustness and ability to rank feature importance. The algorithm assigns an importance score to each feature based on its contribution to the classification task. Features are then ranked in descending order of importance, and a cumulative selection strategy is applied, retaining only those features that collectively account for 80% of the total variance in the dataset. This threshold balances the trade-off between preserving discriminative power and reducing unnecessary complexity, ensuring that the model remains efficient.

4.4. TabNet for dysgraphia classification

Given the tabular nature of the dataset, this study employs TabNet [12], a deep learning architecture specifically designed for structured data. It leverages a sequential attention mechanism that progressively refines feature representations at each decision-making step, allowing the model to focus on the most relevant input features while maintaining robustness against irrelevant noise. A key advantage of TabNet is its built-in interpretability, achieved through attention masks that highlight the most influential features in each prediction. The model consists of a Feature Transformer, which extracts high-level representations from the initial features to capture complex interactions, and an Attentive Transformer, which employs the Sparsemax function to dynamically generate feature-wise attention masks, ensuring that only the most informative features are utilized while reducing redundancy. The decision-making process follows a sequential refinement strategy, where residual information from previously unused features is leveraged at each step to enhance prediction accuracy.

Preferring TabNet for dysgraphia classification over CNNs and LSTMs based on the original time-series handwriting data is primarily due to the relatively small dataset size (120 participants). Deep learning models like CNNs and LSTMs typically require large amounts of labeled data to generalize effectively, whereas tabular models like TabNet are inherently more sample-efficient. TabNet's structured learning mechanism allows it to capture complex relationships within the data while minimizing overfitting, making it particularly well-suited for datasets of this scale.

To maximize TabNet's effectiveness, careful tuning of key hyperparameters is necessary. The number of decision steps (N_{steps}) determines

Table 5
Summary of statistical measures applied to robotic features and the total number of extracted features. The robotic features capture kinematic and dynamic properties of handwriting movements, providing biomechanical insights into dysgraphia classification.

Feature	Extracted statistics	Total
Kinematics		
angle_1_velocity	mean, std, max, min, median, quartile_1, quartile_3	7
angle_1_acceleration	mean, std, max, min, median, quartile_1, quartile_3	7
angle_1_jerk	mean, std, max, min, median, quartile_1, quartile_3	7
angle_1_sin	mean, std, max, min, median, quartile_1, quartile_3	7
angle_1_cos	mean, std, max, min, median, quartile_1, quartile_3	7
angle_1_theta	mean, std, max, min, median, quartile_1, quartile_3	7
angle_1_logCurRadius	mean, std, max, min, median, quartile_1, quartile_3	7
angle_3_velocity	mean, std, max, min, median, quartile_1, quartile_3	7
angle_3_acceleration	mean, std, max, min, median, quartile_1, quartile_3	7
angle_3_jerk	mean, std, max, min, median, quartile_1, quartile_3	7
angle_3_sin	mean, std, max, min, median, quartile_1, quartile_3	7
angle_3_cos	mean, std, max, min, median, quartile_1, quartile_3	7
angle_3_theta	mean, std, max, min, median, quartile_1, quartile_3	7
angle_3_logCurRadius	mean, std, max, min, median, quartile_1, quartile_3	7
angle_5_velocity	mean, std, max, min, median, quartile_1, quartile_3	7
angle_5_acceleration	mean, std, max, min, median, quartile_1, quartile_3	7
angle_5_jerk	mean, std, max, min, median, quartile_1, quartile_3	7
angle_5_sin	mean, std, max, min, median, quartile_1, quartile_3	7
angle_5_cos	mean, std, max, min, median, quartile_1, quartile_3	7
angle_5_theta	mean, std, max, min, median, quartile_1, quartile_3	7
angle_5_logCurRadius	mean, std, max, min, median, quartile_1, quartile_3	7
Dynamics		
torque_1_integral_1	mean, std, max, min, median, quartile_1, quartile_3	7
torque_1_integral_2	mean, std, max, min, median, quartile_1, quartile_3	7
torque_1_integral_3	mean, std, max, min, median, quartile_1, quartile_3	7
torque_2_integral_1	mean, std, max, min, median, quartile_1, quartile_3	7
torque_2_integral_2	mean, std, max, min, median, quartile_1, quartile_3	7
torque_2_integral_3	mean, std, max, min, median, quartile_1, quartile_3	7
torque_3_integral_1	mean, std, max, min, median, quartile_1, quartile_3	7
torque_3_integral_2	mean, std, max, min, median, quartile_1, quartile_3	7
torque_3_integral_3	mean, std, max, min, median, quartile_1, quartile_3	7
torque_4_integral_1	mean, std, max, min, median, quartile_1, quartile_3	7
torque_4_integral_2	mean, std, max, min, median, quartile_1, quartile_3	7
torque_4_integral_3	mean, std, max, min, median, quartile_1, quartile_3	7
	Total Features	**231**

how many times the model iteratively refines its feature selection process. A higher number of steps allows TabNet to explore feature interactions more thoroughly, leading to improved learning. However, increasing N_{steps} also raises computational complexity and training time, making it crucial to find an optimal balance.

The dimensions of the decision and attentive features, denoted as (N_d, N_a), define the size of the latent representations used during feature transformation and attention selection. The parameter N_d specifies the number of neurons in the decision layer, which is responsible for encoding selected features and passing them to subsequent steps. Meanwhile, N_a represents the number of neurons in the Attentive Transformer, which determines the most relevant features at each decision step. Choosing appropriate values for N_d and N_a is critical to prevent overfitting while maintaining sufficient model capacity to capture complex patterns in the data.

Lastly, the regularization parameter (γ) modulates the contribution of each decision step to the final prediction. A lower γ allows some steps to dominate the prediction process, while a higher γ forces a more uniform distribution of attention across decision steps. This encourages the model to consider a broader set of features instead of focusing excessively on a small subset, thus improving generalization and robustness.

Two separate TabNet models were trained independently to assess the relative importance of different handwriting descriptors. The first model focused exclusively on traditional features, based on traditional handwriting descriptors, while the second was trained on robotic features, which exploit the kinematic and dynamic characteristics derived from robotic arm simulations. By treating these feature sets separately, the analysis aimed to uncover distinct patterns within each domain while ensuring that their respective contributions to dysgraphia classification were thoroughly evaluated. Each model produced an independent probability score, denoted as δ_1 for traditional features and δ_2 for robotic features.

TabNet computes class probabilities for each instance, distinguishing between dysgraphic and non-dysgraphic subjects. The probability computation follows the Softmax function, which guarantees that all predicted probabilities sum to one. The mathematical formulation of this process is given by:

$$\boldsymbol{p}(x) = \text{Softmax}\left(\sum_{t=1}^{N_{\text{steps}}} \gamma_t \cdot \mathbf{D}t(x)\right) \quad \text{Softmax}(z_i) = \frac{\exp(z_i)}{\sum_j \exp(z_j)}$$

where $\mathbf{D}_t(x)$ denotes the output of the decision layer at step t, encoding the logits for all classes based on the selected features. The Softmax function then normalizes these logits into probability values, producing the respective classification scores δ_1 and δ_2. These probability scores are independent assessments derived from the two feature sets and will be fused in the final score aggregation step to enhance classification performance.

4.5. Tanh-based fusion

The final step of the proposed pipeline integrates the classification outputs from the two independently trained TabNet models using a Tanh-based score fusion mechanism. This approach aims to enhance overall classification performance by effectively combining the respective scores δ_1 and δ_2, which correspond to predictions based on traditional and robotic features. Score normalization techniques, such as Tanh-based fusion, have been widely used in multimodal biometric systems to improve the robustness of classification decisions by aligning the score distributions from different sources [33]. Through this method, the fusion strategy ensures that both feature sets contribute effectively to the final decision-making process, mitigating disparities in score scaling and improving generalization.

The fusion process relies on the hyperbolic tangent function (tanh), which normalizes the individual scores within a bounded range of $[-1, 1]$. This transformation reduces the impact of extreme values, thereby mitigating the influence of outliers and improving the stability of the final decision score. The fusion operation is mathematically expressed as follows:

$$\delta = \alpha \cdot \tanh(\delta_1) + (1 - \alpha) \cdot \tanh(\delta_2)$$

where δ represents the final fused score, while δ_1 and δ_2 denote the individual probability scores obtained from the respective models. The fusion process is weighted by the parameter α, which determines the relative contribution of each model to the final decision. The parameter α is constrained within the range $0 < \alpha < 1$, allowing flexibility in emphasizing either traditional or robotic features depending on their effectiveness in classification.

5. Experiments

This section outlines the experimental framework designed to assess the proposed method for dysgraphia classification and presents the results obtained.

5.1. Setting

Before training, hyperparameter tuning was performed to determine the optimal configuration for each model. TabNet requires carefully selecting four primary hyperparameters: N_{steps}, N_d, N_a, and γ. The first three parameters ($N_{steps} = 5$, $N_d = 32$, $N_a = 32$), were kept consistent across both models. However, the γ parameter, which controls feature

Table 6
Optimized α values for the Tanh-based score fusion across handwriting tasks.

	l	*l (fast)*	*le*	*le (fast)*	*leto*	*lamoken*	*hračkárstvo*	Sentence	All
α	0.5	0.5	0.15	0.375	0.6	0.775	0.325	0.575	0.475

sparsity, was adjusted to better suit the complexity of each feature set. Specifically, the model trained on robotic features used a slightly higher γ (1.7) compared to the model trained on traditional features (1.5), reflecting the higher complexity of the biomechanical data and the need for enhanced feature selection.

The training followed a stratified 10-fold cross-validation strategy, preserving class distribution across training and validation splits. To prevent overoptimistic results, feature normalization and selection were applied exclusively to the training folds. Model training incorporated early stopping and dynamic learning rate adjustment to enhance generalization and mitigate overfitting. Specifically, the model was trained using the Adam optimizer with an initial learning rate of 0.001, which was dynamically adjusted based on performance. The evaluation metric used for optimization was accuracy, with a maximum of 100 epochs per training run. Early stopping was implemented with a patience of 15 epochs to prevent unnecessary overfitting. A batch size of 32 was used, along with a virtual batch size of 16 to stabilize batch normalization during training.

To further validate model robustness and assess generalizability across subjects, a leave-*p*-out cross-validation was also conducted with $p = 3$. This protocol mimics realistic deployment conditions by training on nearly the entire dataset while testing on previously unseen participants.

After training the two models independently, the Tanh-based score fusion strategy was applied to optimally combine the classification scores (δ_1, δ_2) from the traditional and robotic features. To optimize α, a grid search strategy was applied, varying the parameter within the range $[0.025, 1)$ with a step size of 0.025. Table 6 presents the optimal α values for each handwriting task.

Interestingly, tasks such as writing the syllable *le* or the pseudoword *hračkárstvo* rely more on traditional features, with relatively low α values (0.15 and 0.325, respectively). This suggests that these tasks may involve either highly automatized motor patterns or consistent visual-motor sequences that are well captured by traditional kinematic and pressure-based descriptors. In contrast, more variable and biomechanically demanding tasks, such as writing the pseudoword *lamoken* or the full sentence, show higher α values (0.775 and 0.575, respectively), indicating a greater contribution from robotic-derived features. These tasks likely require complex joint coordination and increased motor planning effort, exposing subtleties in joint torque and angular dynamics not captured by standard features. The *all* task, which concatenates all writing samples, exhibits a relatively balanced fusion parameter ($\alpha = 0.475$), reflecting the complementary nature of both feature types across diverse task conditions.

In addition to accuracy, several evaluation metrics, including sensitivity, specificity, precision, and the Receiver Operating Characteristic (ROC) curve, were used to comprehensively assess model performance.

5.2. Results

The evaluation follows a structured approach. First, an ablation study assesses the impact of normalization and feature selection steps on model performance. Next, models trained separately on traditional and robotic features are compared against the combined model to highlight the benefits of feature fusion. Finally, the proposed approach is benchmarked against state-of-the-art methods, considering both online and offline handwriting data.

5.2.1. Ablation study

An ablation study was conducted across all tasks to quantify the impact of normalization and feature selection. Four configurations were evaluated: the baseline model, trained on raw data without preprocessing; a model trained after applying feature normalization; a model trained using only the most informative features, selected before training; and the final model, which incorporates normalization and feature selection.

The results (Table 7) indicate that the final model consistently outperforms the other configurations across all tasks. This underscores the importance of preprocessing, particularly the combination of normalization and feature selection, in enhancing robustness and generalization. Notably, despite TabNet's built-in feature selection mechanism, applying an additional feature selection step before classification further improved performance.

5.2.2. Impact of feature sets

The impact of different feature sets was analyzed by comparing models trained on traditional features, robotic features, and their combined representation. The results (Table 8), reported with corresponding 95% confidence intervals, show that the combined model achieves the highest accuracy across all tasks, particularly in complex ones like the sentence (90.8%) and *all* (91.7%). The fusion of traditional and robotic features enhances the model's ability to capture both fundamental handwriting traits and biomechanical complexities, leading to an accuracy improvement of approximately 3%–4% in the *all* task compared to the best-performing individual feature set.

In simpler tasks, such as *l* and *le*, the accuracy differences between models are less pronounced, suggesting that either feature set alone is sufficient for these cases. However, the synergy between traditional and robotic features proves crucial in more demanding tasks, validating the effectiveness of the feature fusion strategy.

Beyond accuracy, the models were evaluated in terms of specificity, sensitivity, and precision. The combined model consistently achieves the highest sensitivity across all tasks except *hračkárstvo*, where it performs only 1% below the robotic model. However, the results reveal a more nuanced trend for specificity and precision. The combined model outperforms individual feature models in five out of nine tasks. However, robotic features achieve the highest specificity and precision in *l* (fast), *le*, and *le* (fast), while the traditional features perform best in *leto*. This suggests that, although the combined model delivers superior overall performance, feature fusion does not always yield improvements in every metric. Nevertheless, it provides a well-balanced trade-off across tasks, achieving the best specificity (95.2%), sensitivity (87.3%), and precision (95.2%) in the *all* task, demonstrating its robustness in correctly identifying both positive and negative cases.

To further assess the robustness and generalizability of the proposed model, we conducted an additional leave-p-out cross-validation with $p = 3$. The results, reported in Table 9, include accuracy, specificity, sensitivity, and precision with corresponding 95% confidence intervals. While performance is slightly reduced compared to the original evaluation protocol, the overall trends remain consistent. These findings suggest that the model does not overfit the training data and is capable of generalizing effectively across diverse handwriting tasks and user-specific motor patterns.

To complement these findings, we also analyzed ROC curves across all tasks (Fig. 4). The combined model almost always outperforms the individual models, with its ROC curves positioned closer to the upper-left corner, confirming its superior discriminative power.

Finally, to qualitatively assess the complementarity between traditional and robotic features, we computed the pairwise correlation matrix between the top-ranked features selected from each group. As shown in Fig. 5, conventional features tend to be strongly correlated with one another, likely due to their shared origin from surface-level handwriting signals. In contrast, robotic features exhibit low mutual

Table 7
Ablation study evaluating the impact of feature normalization and selection on classification performance across tasks. The table reports average accuracy (%) with standard deviations.

Configuration	*l*	*l (fast)*	*le*	*le (fast)*	*leto*	*lamoken*	*hračkárstvo*	Sentence	All
TabNet (w/o norm, w/o sel)	87.5 ± 9	85.0 ± 7	82.5 ± 7	83.3 ± 8	83.3 ± 9	85.8 ± 8	**86.7 ± 7**	87.5 ± 4	88.3 ± 7
TabNet (w/ norm, w/o sel)	88.3 ± 8	85.0 ± 8	82.5 ± 10	87.5 ± 7	85.0 ± 8	80.0 ± 7	85.0 ± 9	83.3 ± 10	86.7 ± 7
TabNet (w/o norm, w/ sel)	80.8 ± 7	**86.7 ± 7**	80.0 ± 7	83.3 ± 8	85.8 ± 7	80.0 ± 9	83.3 ± 8	80.8 ± 4	83.3 ± 5
TabNet (w/ norm, w/ sel)	**89.2 ± 7**	83.3 ± 6	**86.7 ± 8**	**88.3 ± 8**	**88.3 ± 8**	**89.2 ± 6**	85.8 ± 6	**90.8 ± 7**	**91.7 ± 5**

Table 8
Performance comparison of models trained on traditional, robotic, and combined feature sets, evaluated using stratified 10-fold cross-validation. Results are reported as mean values with 95% confidence intervals.

Feature Set	Metric	*l*	*l (fast)*	*le*	*le (fast)*	*leto*	*lamoken*	*hračkárstvo*	Sentence	All
Traditional	Acc.	86.7 [79.9, 92.8]	76.7 [72.8, 83.7]	77.5 [73.2, 81.7]	77.5 [71.8, 84.9]	81.7 [79.6, 88.9]	87.5 [79.9, 96.0]	84.2 [78.4, 88.2]	82.5 [76.9, 86.4]	88.3 [80.9, 93.7]
	Spec.	90.7 [78.2, 97.5]	78.1 [71.4, 86.3]	83.6 [75.5, 89.8]	90.5 [79.7, 98.7]	**98.3 [91.4, 100.0]**	88.8 [81.4, 96.4]	91.0 [83.5, 97.7]	87.9 [80.4, 94.8]	93.6 [84.8, 100.0]
	Sens.	83.0 [75.5, 93.7]	76.3 [72.9, 85.8]	69.0 [65.0, 84.9]	62.7 [55.2, 75.8]	63.7 [56.0, 78.6]	86.3 [79.0, 94.6]	77.7 [73.5, 84.3]	77.3 [71.2, 81.7]	82.0 [65.3, 90.3]
	Pre.	89.4 [78.0, 98.0]	76.5 [71.0, 82.6]	86.5 [80.6, 93.2]	87.7 [77.6, 97.9]	**98.0 [90.8, 100.0]**	88.5 [80.6, 89.7]	91.3 [84.6, 97.8]	88.4 [80.3, 95.1]	93.0 [84.7, 99.2]
Robotic	Acc.	85.8 [78.5, 91.2]	82.5 [80.9, 89.1]	86.7 [78.4, 90.6]	85.8 [80.2, 95.2]	80.8 [77.3, 87.6]	80.0 [75.9, 85.7]	78.3 [74.9, 87.0]	81.7 [76.9, 89.8]	88.3 [77.4, 92.3]
	Spec.	92.4 [77.7, 97.3]	**87.9 [77.3, 94.6]**	**95.2 [80.3, 100.0]**	**96.7 [85.7, 100.0]**	86.9 [79.7, 99.3]	83.8 [76.0, 92.1]	75.7 [70.7, 87.8]	90.2 [73.0, 100.0]	91.9 [78.8, 100.0]
	Sens.	79.3 [74.0, 95.3]	77.7 [67.6, 83.1]	76.7 [67.9, 91.4]	73.3 [69.1, 87.5]	73.7 [65.0, 86.9]	78.3 [64.9, 87.7]	**80.3 [74.6, 86.0]**	72.0 [64.9, 91.2]	84.0 [70.6, 95.4]
	Pre.	92.6 [79.7, 98.6]	**86.6 [78.4, 93.9]**	**94.3 [82.9, 100.0]**	**95.8 [83.0, 100]**	88.6 [80.3, 98.4]	84.9 [75.6, 91.6]	80.3 [75.5, 92.2]	89.0 [80.1, 100.0]	91.7 [85.6, 100.0]
Traditional + Robotic	Acc.	**89.2 [84.5, 96.6]**	**83.3 [79.3, 89.2]**	**86.7 [81.3, 91.6]**	**88.3 [83.7, 94.6]**	**88.3 [79.3, 96.3]**	**89.2 [80.6, 97.3]**	**85.8 [79.8, 91.2]**	**90.8 [82.6, 98.1]**	**91.7 [86.1, 100.0]**
	Spec.	**95.7 [92.3, 100.0]**	82.4 [74.7, 93.2]	93.6 [86.6, 100.0]	90.5 [85.6, 98.9]	95.2 [85.4, 100.0]	**89.0 [82.8, 94.7]**	**92.1 [85.9, 98.9]**	**93.6 [85.3, 100.0]**	**95.2 [91.3, 100.0]**
	Sens.	**83.3 [74.7, 92.1]**	**84.3 [74.7, 95.2]**	**78.7 [73.6, 84.3]**	**85.3 [74.1, 93.4]**	**81.0 [76.3, 88.7]**	**90.0 [80.6, 97.3]**	79.3 [72.6, 87.2]	**87.7 [81.3, 90.6]**	**87.3 [78.8, 97.2]**
	Pre.	**95.5 [84.9, 100.0]**	81.9 [72.2, 92.7]	92.7 [84.5, 100.0]	90.5 [84.3, 96.7]	94.7 [89.1, 100.0]	**89.6 [78.6, 97.8]**	**93.0 [86.1, 100.0]**	**93.8 [87.4, 100.0]**	**95.2 [85.9, 100.0]**

Acc: Accuracy, Spec: Specificity, Sens: Sensitivity, Pre: Precision.

Table 9
Performance of the model trained on combined traditional and robotic features, evaluated using leave-p-out cross-validation with $p = 3$. Results are reported with 95% confidence intervals.

Feature Set	Metric	*l*	*l (fast)*	*le*	*le (fast)*	*leto*	*lamoken*	*hračkárstvo*	Sentence	All
Traditional + Robotic	Acc.	86.0 [80.1, 91.9]	80.5 [75.2, 85.7]	83.0 [77.5, 88.3]	84.4 [79.1, 89.3]	83.7 [77.8, 89.2]	86.3 [80.5, 91.2]	82.2 [76.8, 87.1]	85.3 [80.2, 90.0]	87.4 [82.3, 92.2]
	Spec.	90.2 [86.4, 98.7]	84.7 [77.3, 91.0]	88.7 [83.2, 95.4]	89.3 [82.4, 95.2]	92.1 [85.1, 97.5]	87.8 [80.2, 94.3]	89.2 [81.5, 95.6]	89.6 [82.5, 95.5]	90.2 [84.6, 96.8]
	Sens.	76.5 [67.1, 84.9]	78.3 [69.0, 86.1]	73.0 [64.3, 81.5]	78.4 [69.8, 85.9]	74.9 [66.2, 83.1]	81.7 [73.1, 89.3]	75.6 [67.0, 83.5]	79.8 [71.4, 87.3]	81.5 [73.2, 88.9]
	Pre.	88.1 [78.3, 96.5]	81.0 [70.9, 90.1]	85.0 [75.2, 93.7]	84.1 [74.0, 92.5]	86.2 [76.8, 94.6]	84.6 [74.7, 92.9]	86.3 [76.5, 94.3]	87.2 [77.3, 95.1]	89.0 [80.2, 96.0]

Acc: Accuracy, Spec: Specificity, Sens: Sensitivity, Pre: Precision.

correlation, suggesting they encode diverse aspects of motor behavior. Moreover, the two groups are only weakly correlated with each other, reinforcing the idea that robotic features capture complementary biomechanical information that is not accessible through traditional digitizing tablets.

5.2.3. Comparison with the state-of-the-art

The final evaluation compares the proposed model with state-of-the-art methods on the same dataset, considering online and offline handwriting analysis approaches (Table 10).

The proposed approach significantly outperforms previous methods from Drotár et al. [9] and Kunhoth et al. [13] across all tasks and metrics for online handwriting models. The accuracy improvement is particularly evident in the *all* task, where our model achieves 91.7 ± 5 compared to 80.8 in Kunhoth et al. The integration of traditional and robotic features, which previous works did not consider, played a crucial role in this improvement. Furthermore, adopting TabNet, with its ability to perform an internal dynamic feature selection, proved advantageous compared to AdaBoost-based models, which rely on fixed feature representations.

In the case of offline handwriting models, the proposed approach demonstrates superior or competitive performance compared to CNN-based approaches. The DenseNet+SVM method by Kunhoth et al. [23] achieves strong performance in specific tasks, such as *leto* (91.7±4), *lamoken* (90.0 ± 4), and *hračkárstvo* (87.1 ± 2). However, our model remains within a 3%–4% accuracy difference in these cases, surpassing it in the sentence and *all* tasks. This result highlights the advantage of leveraging temporal, kinematic, and dynamic handwriting features from online data, which offline models fail to capture, as they are limited to spatial characteristics extracted from static handwriting images.

Beyond accuracy, the proposed model achieves higher precision across all tasks, demonstrating its ability to generalize effectively across simple and complex scenarios.

6. Conclusions

This study introduced an innovative handwriting analysis framework that integrates robotic features with traditional online features to achieve a deeper understanding of patterns in dysgraphic and non dysgraphic subjects. By training TabNet models separately on each feature set and employing a Tanh-based score fusion mechanism, the proposed approach dynamically balances model outputs to optimize classification performance across tasks. The experimental results confirm a significant improvement in terms of accuracy over state-of-the-art methods, demonstrating the effectiveness of this methodology in analyzing handwriting behavior. These findings represent a step forward in leveraging AI-driven techniques for dysgraphia classification, offering a robust framework for real-world applications.

Nonetheless, this study has some limitations. First, the evaluation was conducted using a single publicly available dataset, which, although widely used, may not capture the full variability present in real-world handwriting scenarios. Second, while the fusion of robotic and traditional features demonstrated strong performance, the interpretability of the robotic features in clinical terms warrants further investigation. Finally, the current pipeline was tested in an offline setting; additional studies are needed to assess its responsiveness and usability in real-time contexts.

Future research directions could explore the real-time applicability of this framework, adapting it to streaming handwriting data environments where continuous user feedback is provided. Such an extension would enhance usability in educational and clinical settings, enabling immediate intervention for handwriting difficulties. Integrating multimodal data — such as audio, video, and neurophysiological signals — could enrich handwriting analysis, improving diagnostic precision and offering deeper behavioral insights. Another promising avenue involves deploying this system on mobile and tablet-based platforms, facilitating widespread accessibility and practical adoption for dysgraphia screening and handwriting training. These advancements would pave the

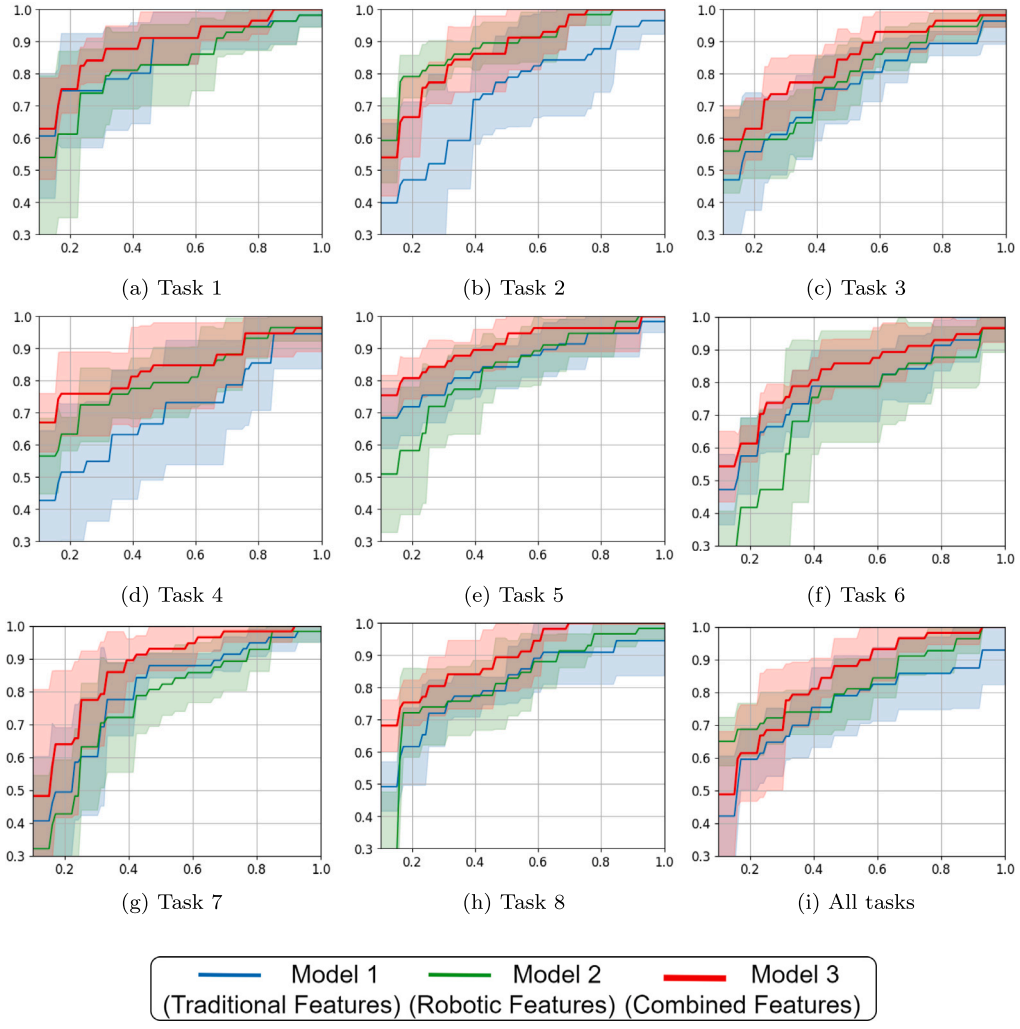

Fig. 4. ROC curves. Each subfigure corresponds to a specific task, while the last one (bottom right) represents the aggregated results across all tasks. The ROC curve illustrates the performance of a classification model at various threshold settings. The x-axis represents the false positive rate, while the y-axis represents the true positive rate.

Table 10
Comparison of accuracy, specificity, sensitivity, and precision across online and offline state-of-the-art methods. For a fairer comparison, standard deviations are reported alongside average scores.

Ref.	Model	Metric	*l*	*l (fast)*	*le*	*le (fast)*	*leto*	*lamoken*	*hračkárstvo*	Sentence	All
[9]	AdaBoost	Acc.	68.2 ± 2	60.4 ± 3	67.6 ± 2	61.0 ± 3	65.8 ± 2	66.4 ± 3	76.0 ± 2	64.4 ± 2	79.5 ± 3
		Spec.	64.5 ± 4	64.6 ± 4	73.5 ± 4	65.6 ± 4	69.5 ± 5	73.3 ± 5	79.5 ± 3	66.1 ± 3	76.7 ± 2
		Sens.	55.8 ± 4	55.9 ± 4	61.4 ± 3	55.6 ± 3	61.5 ± 4	58.6 ± 4	73.1 ± 3	62.9 ± 2	79.7 ± 5
		Pre.	–	–	–	–	–	–	–	–	–
[13]	AdaBoost	Acc.	–	–	–	–	–	–	–	–	80.8
		Spec.	–	–	–	–	–	–	–	–	–
		Sens.	–	–	–	–	–	–	–	–	78.5
		Pre.	–	–	–	–	–	–	–	–	83.3
[22]	CNN	Acc.	–	–	–	–	–	–	–	–	79.7 ± 3
		Spec.	–	–	–	–	–	–	–	–	76.7 ± 2
		Sens.	–	–	–	–	–	–	–	–	79.5 ± 3
		Pre.	–	–	–	–	–	–	–	–	–
[23]	DenseNet + SVM	Acc.	–	–	–	–	**91.7 ± 4**	**90.0 ± 4**	**87.1 ± 2**	88.7 ± 4	–
		Spec.	–	–	–	–	–	–	–	–	–
		Sens.	–	–	–	–	**90.4 ± 7**	88.2 ± 6	**84.6 ± 4**	86.8 ± 7	–
		Pre.	–	–	–	–	91.7 ± 4	90.0 ± 4	87.1 ± 2	88.7 ± 4	–
This work	TabNet	Acc.	**89.2 ±7**	**83.3 ±6**	**86.7 ± 8**	**88.3 ± 8**	88.3 ± 8	89.2 ± 6	85.8 ± 6	**90.8 ± 7**	**91.7 ±5**
		Spec.	**95.7 ± 9**	**82.4 ± 9**	**93.6 ± 8**	**90.5 ± 15**	**95.2 ± 7**	**89.0 ± 12**	**92.1 ± 13**	**93.6 ± 11**	**95.2 ± 7**
		Sens.	**83.3 ± 17**	**84.3 ± 12**	**78.7 ± 13**	**85.3± 12**	81.0 ± 14	**90.0 ± 11**	79.3 ± 17	**87.7 ± 8**	**87.3 ± 12**
		Pre.	**95.5 ± 9**	**81.9± 7**	**92.7 ± 9**	**90.5 ± 12**	**94.7 ±8**	**90.1 ±11**	**93.0 ± 11**	**93.8 ± 10**	**95.2 ± 7**

Acc: Accuracy, Spec: Specificity, Sens: Sensitivity, Pre: Precision.

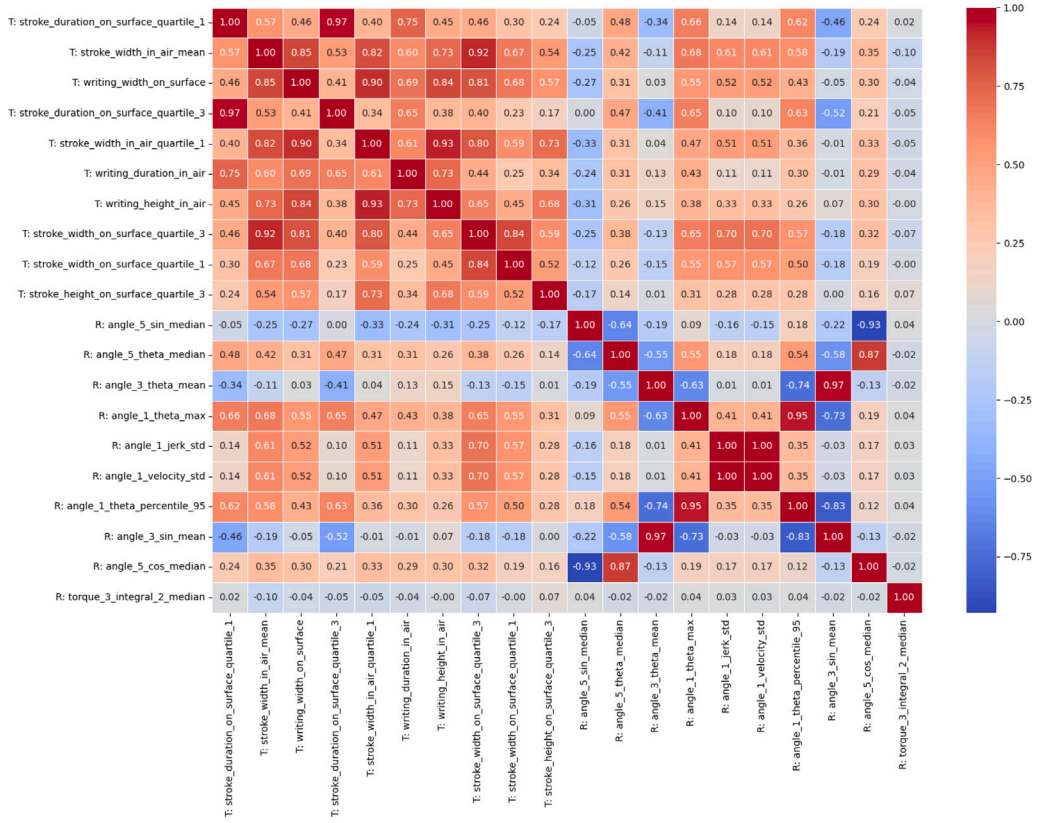

Fig. 5. Correlation matrix between the top 10 traditional features (T) and the top 10 robotic features (R), selected via feature importance ranking.

way for next-generation handwriting analysis tools, fostering academic research and real-world impact.

CRediT authorship contribution statement

Francesco Brescia: Writing – original draft, Investigation, Conceptualization. **Belén Esther Alemán Santana:** Data curation, Conceptualization. **Moises Diaz:** Writing – review & editing, Supervision, Conceptualization. **Gennaro Vessio:** Writing – review & editing, Supervision, Conceptualization, Writing – original draft. **Miguel Angel Ferrer:** Writing – review & editing, Conceptualization. **Giovanna Castellano:** Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was partly supported by the PID2023-146620OB-I00, funded by MICIU/AEI 10.13039/501100011033 and the European Union's ERDF program, and partly by the CajaCanaria and la Caixa (2023DIG05). While preparing this work, the authors utilized ChatGPT and Grammarly to enhance language clarity and readability. The authors, who take full responsibility for the final version of the manuscript, carefully reviewed and refined all content generated by these tools.

Data availability

The authors do not have permission to share data.

References

[1] T.H. Falk, C. Tam, H. Schellnus, T. Chau, On the development of a computer-based handwriting assessment tool to objectively quantify handwriting proficiency in children, Comput. Methods Programs Biomed. 104 (3) (2011) e102–e111.

[2] K.H. James, L. Engelhardt, The effects of handwriting experience on functional brain development in pre-literate children, Trends Neurosci. Educ. 1 (1) (2012) 32–42.

[3] P.A. Mueller, D.M. Oppenheimer, The pen is mightier than the keyboard: Advantages of longhand over laptop note taking, Psychol. Sci. 25 (2014) 1159–1168, http://dx.doi.org/10.1177/0956797614524581.

[4] M. Moetesum, M. Diaz, U. Masroor, I. Siddiqi, G. Vessio, A survey of visual and procedural handwriting analysis for neuropsychological assessment, Neural Comput. Appl. 34 (12) (2022) 9561–9578.

[5] T.J. Roberts, D.J. Williams, Understanding dysgraphia: A neurological perspective, J. Learn. Disabil. 42 (2009) 469–477, http://dx.doi.org/10.1177/0022219409332227.

[6] M.-H. Tseng, J.D. Li, Motor deficits and handwriting difficulties in children with dysgraphia, J. Child Neurol. 24 (2009) 515–522, http://dx.doi.org/10.1177/0883073809334849.

[7] K. Rayner, D.R. Liversedge, Dysgraphia and its impact on written expression: A cognitive neuroscience perspective, Cogn. Psychol. 64 (2012) 241–268, http://dx.doi.org/10.1016/j.cogpsych.2012.02.001.

[8] J. Mekyska, M. Faundez-Zanuy, Z. Mzourek, Z. Galaz, Z. Smekal, S. Rosenblum, Identification and rating of developmental dysgraphia by handwriting analysis, IEEE Trans. Human-Machine Syst. 47 (2) (2016) 235–248.

[9] P. Drotár, M. Dobeš, Dysgraphia detection through machine learning, Sci. Rep. 10 (1) (2020) 21541.

[10] M. Diaz, M.A. Ferrer, J.J. Quintana, Anthropomorphic features for on-line signatures, IEEE Trans. Pattern Anal. Mach. Intell. 41 (12) (2018) 2807–2819.

[11] M. Diaz, M.A. Ferrer, J.J. Quintana, A. Wolniakowski, R. Trochimczuk, K. Miatliuk, G. Castellano, G. Vessio, Neural network modelling of kinematic and dynamic features for signature verification, Pattern Recognit. Lett. 187 (2025) 130–136.

[12] S.Ö. Arik, T. Pfister, TabNet: Attentive interpretable tabular learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, (no. 8) 2021, pp. 6679–6687.

[13] J. Kunhoth, S. Al Maadeed, M. Saleh, Y. Akbari, Exploration and analysis of On-Surface and In-Air handwriting attributes to improve dysgraphia disorder

diagnosis in children based on machine learning methods, Biomed. Signal Process. Control. 83 (2023) 104715.
[14] S. Rosenblum, G. Dror, Identifying developmental dysgraphia characteristics utilizing handwriting classification methods, IEEE Trans. Human-Machine Syst. 47 (2) (2016) 293–298.
[15] T. Asselborn, T. Gargot, Ł. Kidziński, W. Johal, D. Cohen, C. Jolly, P. Dillenbourg, Automated human-level diagnosis of dysgraphia using a consumer tablet, NPJ Digit. Med. 1 (1) (2018) 42.
[16] V. Zvoncak, J. Mucha, Z. Galaz, J. Mekyska, K. Safarova, M. Faundez-Zanuy, Z. Smekal, Fractional order derivatives evaluation in computerized assessment of handwriting difficulties in school-aged children, in: 2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops, ICUMT, IEEE, 2019, pp. 1–6.
[17] Z. Dankovičová, J. Hurtuk, P. Fecil'ak, Evaluation of digitalized handwriting for dysgraphia detection using random forest classification method, in: 2019 IEEE 17th International Symposium on Intelligent Systems and Informatics, SISY, IEEE, 2019, pp. 000149–000154.
[18] P.W. Samodro, S.W. Sihwi, et al., Backpropagation Implementation to Classify Dysgraphia In Children, in: 2019 International Conference of Artificial Intelligence and Information Technology, ICAIIT, IEEE, 2019, pp. 437–442.
[19] R. Kariyawasam, M. Nadeeshani, T. Hamid, I. Subasinghe, P. Samarasinghe, P. Ratnayake, Pubudu: Deep learning based screening and intervention of dyslexia, dysgraphia and dyscalculia, in: 2019 14th Conference on Industrial and Information Systems, ICIIS, IEEE, 2019, pp. 476–481.
[20] 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.
[21] F. Masood, W.U. Khan, K. Ullah, A. Khan, F.H. Alghamedy, H. Aljuaid, A hybrid CNN-LSTM random forest model for dysgraphia classification from hand-written characters with uniform/normal distribution, Appl. Sci. 13 (7) (2023) 4275.
[22] J. Skunda, B. Nerusil, J. Polec, Method for Dysgraphia Disorder Detection Using Convolutional Neural Network, Václav Skala-UNION Agency, 2022.
[23] J. Kunhoth, S. Al Maadeed, M. Saleh, Y. Akbari, CNN feature and classifier fusion on novel transformed image dataset for dysgraphia diagnosis in children, Expert Syst. Appl. 231 (2023) 120740.
[24] H.H. Nguyen, M.B. Blaschko, S. Saarakkala, A. Tiulpin, Clinically-inspired multi-agent transformers for disease trajectory forecasting from multimodal data, IEEE Trans. Med. Imaging 43 (1) (2023) 529–541.
[25] H.-Y. Zhou, Y. Yu, C. Wang, S. Zhang, Y. Gao, J. Pan, J. Shao, G. Lu, K. Zhang, W. Li, A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics, Nat. Biomed. Eng. 7 (6) (2023) 743–755.
[26] S.-Y. Lu, Y.-D. Zhang, Y.-D. Yao, A regularized transformer with adaptive token fusion for Alzheimer's disease diagnosis in brain magnetic resonance images, Eng. Appl. Artif. Intell. 155 (2025) 111058.
[27] J. Denavit, R.S. Hartenberg, A Kinematic Notation for Lower-Pair Mechanisms Based on Matrices, American Society of Mechanical Engineers, 1955.
[28] J.C. Lagarias, J.A. Reeds, M.H. Wright, P.E. Wright, Convergence properties of the Nelder–Mead simplex method in low dimensions, SIAM J. Optim. 9 (1) (1998) 112–147.
[29] J.J. Craig, Introduction to Robotics: Mechanics and Control, vol. 3, Pearson Prentice Hall Upper Saddle River, 2005.
[30] M. Faundez-Zanuy, O. Brotons-Rufes, C. Paul-Recarens, R. Plamondon, On handwriting pressure normalization for interoperability of different acquisition stylus, IEEE Access 9 (2021) 18443–18453.
[31] J. Jiang, S. Lai, L. Jin, Y. Zhu, DsDTW: Local representation learning with deep soft-DTW for dynamic signature verification, IEEE Trans. Inf. Forensics Secur. 17 (2022) 2198–2212.
[32] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[33] A. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, Pattern Recognit. 38 (12) (2005) 2270–2285.