Computer Methods and Programs in Biomedicine 247 (2024) 108066

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

LSTM-CNN: An efficient diagnostic network for Parkinson's disease utilizing dynamic handwriting analysis

Xuechao Wang [a,*,1], Junqing Huang [a,1], Marianna Chatzakou [a], Kadri Medijainen [c], Aaro Toomela [d], Sven Nõmm [b], Michael Ruzhansky [a,e]

[a] Department of Mathematics: Analysis, Logic and Discrete Mathematics, Ghent University, Ghent, Belgium
[b] Department of Software Science, Faculty of Information Technology, Tallinn University of Technology, Akadeemia tee 15 a, 12618, Tallinn, Estonia
[c] Institute of Sport Sciences and Physiotherapy, University of Tartu, Puusepa 8, Tartu 51014, Estonia
[d] School of Natural Sciences and Health, Tallinn University, Narva mnt. 25, 10120, Tallinn, Estonia
[e] School of Mathematical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, United Kingdom

ARTICLE INFO

Keywords:
Parkinson's disease
Dynamic handwriting analysis
Long short-term memory
Convolutional neural network
Real-time diagnosis

ABSTRACT

Background and objectives:: Dynamic handwriting analysis, due to its noninvasive and readily accessible nature, has emerged as a vital adjunctive method for the early diagnosis of Parkinson's disease (PD). An essential step involves analysing subtle variations in signals to quantify PD dysgraphia. Although previous studies have explored extracting features from the overall signal, they may ignore the potential importance of local signal segments. In this study, we propose a lightweight network architecture to analyse dynamic handwriting signal segments of patients and present visual diagnostic results, providing an efficient diagnostic method.
Methods:: To analyse subtle variations in handwriting, we investigate time-dependent patterns in local representation of handwriting signals. Specifically, we segment the handwriting signal into fixed-length sequential segments and design a compact one-dimensional (1D) hybrid network to extract discriminative temporal features for classifying each local segment. Finally, the category of the handwriting signal is fully diagnosed through a majority voting scheme.
Results:: The proposed method achieves impressive diagnostic performance on the new DraWritePD dataset (with an accuracy of 96.2%, sensitivity of 94.5% and specificity of 97.3%) and the well-established PaHaW dataset (with an accuracy of 90.7%, sensitivity of 94.3% and specificity of 87.5%). Moreover, the network architecture stands out for its excellent lightweight design, occupying a mere 0.084M parameters, with only 0.59M floating-point operations. It also exhibits nearly real-time CPU inference performance, with the inference time for a single handwriting signal ranging from 0.106 to 0.220 s.
Conclusions:: We present a series of experiments with extensive analysis, which systematically demonstrate the effectiveness and efficiency of the proposed method in quantifying dysgraphia for a precise diagnosis of PD.

1. Introduction

Parkinson's disease (PD) is one of the most widespread and disabling neurodegenerative disorders. Symptoms typically appear after approximately 70% of dopamine-producing cells cease to function normally, resulting in a decrease in functional, cognitive and behavioural abilities [1]. Currently, there is no cure, but if PD is diagnosed early, the progression of the disease can be significantly slowed with appropriate treatment methods [2,3]. Typical early motor symptoms include resting tremors, rigidity, postural instability, and bradykinesia (i.e., slowness of spontaneous movement [4]), where resting tremors are one of the most common symptoms [5]. These motor symptoms, together with nonmotor symptoms, have many adverse effects on the patient's quality of life, family relationships, and social functioning [6], while they can also increase the risk of further health complications. This can place a significant economic burden on the individual and society.

Handwriting is a common but complex human activity in a variety of leisure and professional settings, which requires fine dexterity skills and

* Corresponding author.
E-mail address: xuechao.wang@ugent.be (X. Wang).
[1] Both authors contributed equally.

https://doi.org/10.1016/j.cmpb.2024.108066
Received 4 December 2023; Received in revised form 30 January 2024; Accepted 7 February 2024
Available online 9 February 2024
0169-2607/© 2024 Elsevier B.V. All rights reserved.

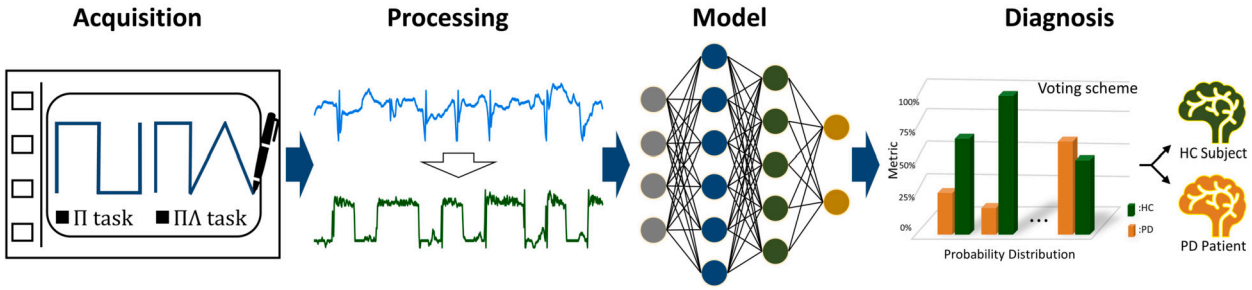


Fig. 1. Overview of the proposed framework for Parkinson's disease diagnosis.

involves an intricate blend of cognitive, sensory, and perceptual-motor components [7]. Changes in handwriting have been well documented to be promising biomarkers for the early diagnosis of PD [8,9], and once diagnosed, allow for later neuroprotective interventions. However, related studies have indicated that the accuracy of clinical diagnosis is relatively low [10], while the cost of diagnosis is quite expensive. Fortunately, a growing body of knowledge provides evidence that it is possible to distinguish between unhealthy and healthy individuals automatically using simple and easy-to-perform handwriting tasks [8, 11]. Therefore, the development of handwriting-based decision support tools is necessary to obtain noninvasive and low-cost solutions that can support current standard clinical assessments.

Dynamic (online) systems using digital tablets [12] or biometric smart pen [13] facilitate the capture of various temporal and spatial sequence signals of handwriting. This enables the quantification of temporal, kinematic, and dynamic manifestations of PD handwriting impairments, which cannot be objectively studied using a classical paper and pen method. Several studies have explored the impact of quantitative analysis of PD dysgraphia [8,11]. They primarily derive various dynamic features through artificial engineering, including conventional kinematic features and less common features based on entropy and signal energy [14–18]. To obtain a "holistic" statistical representation of dynamic features, it is often necessary to calculate single-valued descriptors of the feature vector, such as mean, median, standard deviation, and others. However, compressing a vector of features of an arbitrary length into a single value can ignore crucial local details. Another method of analysing handwriting patterns is to use convolutional neural networks (CNN) to autonomously extract features from two-dimensional (2D) images generated from "holistic" handwriting signals [19–23]. This may be considered robust alternatives to artificial engineering, but only offers a 'holistic' perspective on the handwriting studied. Furthermore, since it is black-box, this method obscures the interpretation of the employed features and their correlation with the accompanying disease. Despite the significant progress in the analysis of online handwriting to assist in diagnosing PD, further improvements in performance and interpretability are expected to improve diagnostic efficiency and facilitate its application in the medical field.

In this paper, we investigate the use of one-dimensional (1D) dynamic handwriting signal segments to extract local handwriting sequence patterns from patients with PD and present intuitive visual diagnosis results. The main idea arises from the fact that patients with PD frequently manifest more distinct handwriting impairments during the local drawing process, such as resting tremors, rigidity, and bradykinesia [4]. Specifically, we propose using a fixed-length sliding window for local sampling of the handwriting signal to retain local time-dependent patterns. Based on this, we design a compact and efficient hybrid model, named LSTM-CNN, that integrates RNN and CNN, taking 1 D handwriting sequence segments as input. Finally, the final result of the diagnosis is determined by combining the inferential diagnosis strategy with a majority voting scheme. The proposed method effectively leverages the sequential nature of the data, explicitly incorporates local temporal information, and offers novel insight into the handwriting process. The key contributions of this work are as follows:

- We propose an efficient method for analyzing handwriting sequence segments to facilitate the early diagnosis of PD.
- A concise sliding window data segmentation technique is introduced to preserve local temporal information.
- A hybrid neural network incorporating LSTM and 1D convolution is designed, demonstrating outstanding performance while maintaining a lightweight structure.
- We investigate three types of handwriting tasks and systematically confirm the superior performance of the proposed method compared to the state-of-the-art methods.

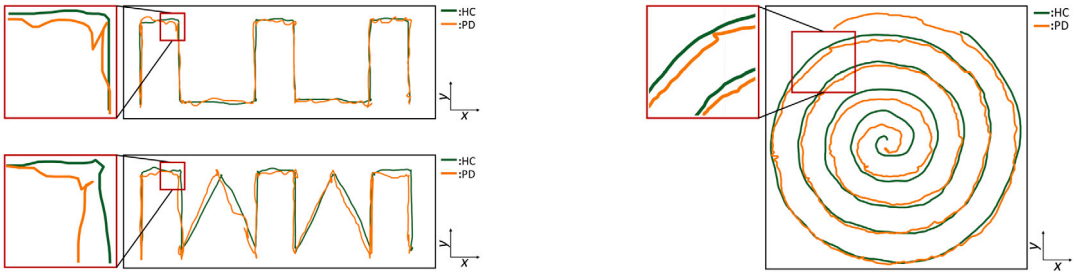
The paper is organised as follows. Section 2 introduces the related work on the diagnosis of PD based on dynamic handwriting signals. Section 3 provides the reader with the necessary information about the data. Data processing and model details are described in Section 4. Section 5 presents the main results of the current studies. Finally, the discussion of the results achieved, the limitations of the proposed methods in Section, and the possible future directions are discussed in Sections 6 and 7.

2. Related work

The application of machine learning techniques in various clinical contexts has gained significant momentum, particularly driving the advancement of automated decision systems for the diagnosis of PD [8,11]. Here we mainly review the related methods based on dynamic handwriting data and elucidate their interconnections and potential advantages.

A series of outstanding works have been extensively reported, which are based on manual feature engineering and aim to extract discriminative features from handwriting signals, subsequently integrating them with traditional machine learning classification models. For example, Drotár et al. [15] investigate a variety of kinematic-based features, and achieve an accuracy 85% on their publicly available Parkinson's Disease Handwriting (PaHaW) data set [15,17]. In a subsequent study [17], Drotár et al. further underscored the significance of novel pressure-related features in the context of the diagnosis of PD. Mucha et al. [24] explore the impact of online handwriting parameterisation based on novel fractional order derivatives on the diagnosis and monitoring of PD dysgraphia. Impedovo [25] combines classical features with new velocity-based features to extend the handcrafted feature set. These methods lead to improved results on the PaHaW dataset. More recently, Valla et al. [18] have explored new derivative-based, angle-type, and integral-like features of the Archimedes spiral graph test, demonstrating their value in dynamic handwriting analysis.

Simultaneously, among well-known contributions based on reconstructing 2D images, the NewHandPD dataset is introduced in [19], which is a set of signals extracted from a smart pen, Pereira et al. convert the dynamic signal into an image and the problem of diagnosing PD is regarded as an image recognition task. This study is one of the first applications of a deep learning-orientated 2 D method to diagnose PD. Subsequently, the work is extended further in [20] and [22]. Specifically, Pereira et al. [20] combine with various CNN configurations and



(a) Π Task and ΠΛ Task (b) Spiral Task

Fig. 2. Comparison of the hand drawings from the Parkinson's disease (PD) patients and healthy control (HC) subjects, where the relative positions of the hand drawings are reconstructed based on the two-dimensional coordinates of successive hand drawing points over time.

majority voting schemes to learn texture-orientated features directly from time series-based images. Afonso et al. [22] propose applying a recurrent graph to map the signal in the image domain and input them into CNNs to learn the appropriate information. Subsequently, the discriminative power of "dynamically enhanced" static handwriting images is investigated in [21]. The authors propose a static representation that embeds dynamic information. Nõmm et al. [23] improve the Archimedes spiral drawing images by controlling the thickness and colour of the spiral drawing according to the kinematics and pressure features, while preserving the original shape of the drawing curve. They achieve impressive performance using the classic AlexNet [26].

More recently, RNN-based models have been preliminary explored [27] to capture temporal information derived from temporal dependencies within handwriting signals [28]. Diaz et al. [29] propose using CNNs to extract features from the original feature set and its derived feature set, followed by classification using bidirectional gated recurrent unit (GRU) [30]. This method yields remarkably high diagnostic performance on the PaHaW and NewHandPD datasets. Although RNN-based networks have not undergone exhaustive exploration in this domain, these studies underscore the potential of 1D sequence-based dynamic data analysis in early PD diagnosis.

3. Materials

Two datasets are employed in this study. The first dataset, DraWritePD (acquired by the authors), is used for system fine-tuning to determine the optimal configuration. The second dataset, PaHaW [15,17], served as an additional testing set to evaluate the performance of the proposed method. As illustrated in Fig. 2, three distinct sets of handwriting tasks are employed to validate the robustness of the proposed system. More details of the two datasets are described as follows.

3.1. DraWritePD

The DraWritePD dataset contains handwriting data from 20 patients with PD, with an average age of 74.1 ± 6.7. A control group of 29 healthy control (HC) subjects, gender and age-matched (average age 74.1 ± 9.1), meeting all the inclusion criteria except the PD diagnosis, are taken as the HC group. Data acquire with an iPad Pro (9.7 inches) equipped with a stylus. The patients with PD and their HC counterparts are asked to complete a series of handwriting tests consisting of 12 drawing and writing tasks. Fig. 2(a) and Fig. 2(b) present the shapes of the Π task and ΠΛ task, respectively. To familiarize subjects with the iPad, they are asked to practice drawing the state line, writing their name, and writing digits from 0 to 9. During each task, the iPad Pro scans the stylus signal at a fixed rate. For each data point, the dynamic sequence signal captures six sets of independent dynamic features, including y-coordinate (mm), x-coordinate (mm), timestamp (s), azimuth (rad), altitude (rad), pressure (arbitrary unit of force applied on the surface). The data acquisition process is conducted under strict privacy law guidelines. The study is approved by the Research Ethics Committee of the University of Tartu (No. 1275T – 9).

3.2. PaHaW

The PaHaW dataset collects handwriting data from 37 PD patients and 38 age and gender-matched HC subjects [15,17]. During the acquisition of PaHaW dataset, each subject is asked to complete handwriting tasks according to the prepared pre-filled template at a comfortable speed. Fig. 2(c) shows the shape of the spiral task. Handwriting signals are recorded using a digitizing tablet overlaid with a blank sheet of paper (signals are recorded using an Intuos 4M pen of frequency 200 Hz). For each data point, the dynamic signal captures seven sets of independent dynamic features, including: y-coordinate, x-coordinate, timestamp, button status, altitude, azimuth, pressure. All variables are converted to the same units as in DraWritePD.

4. Methodology

This section briefly introduces the proposed method for automatically diagnosing PD. Fig. 1 illustrates the schematic workflow of the proposed method, with further details provided in the subsequent subsections.

4.1. Data processing

The handwriting signals in this study undergo the following three stages of data processing to generate a dataset of sequence patches for training: normalization, enhancement, and segmentation. First, we use the min-max normalization to ensure that data across features share a similar scale. Note that we only consider five features for the classification: x- and y-coordinates, azimuth, altitude, and pressure [15,17,25].

The second step involves applying a simple yet effective differencing operation to highlight motion patterns associated with PD. Specifically, we use a first-order forward difference [24] individually on the x- and y-coordinate features to extract positional changes in the handwriting signal between adjacent data points, while keeping the azimuth, altitude, and pressure features unchanged. The main idea emerged from that, tremor or jerk have been observed on meanders, horizontal, and spiral drawing [4,8] Fig. 2, in the zoomed-in patches, reveals that patients with PD introduce more changes in velocity, and their drawing became much more non-fluent. The extraction of relevant features from handwriting signals has always played a critical role in the PD diagnostic [8,11]. This strategy facilitates the full use of the local information on handwriting patterns, in particular that the high sampling rate of the acquisition device provides a precise approximation of the actual pen trajectory. Note that we use zero padding to ensure all five individual dynamic features maintain a uniform size during the process.

Finally, a sequence segmentation technique is applied to preserve local temporal information, where the sequence data is cropped into multiple sequence patches, controlled by two parameters: *window size*

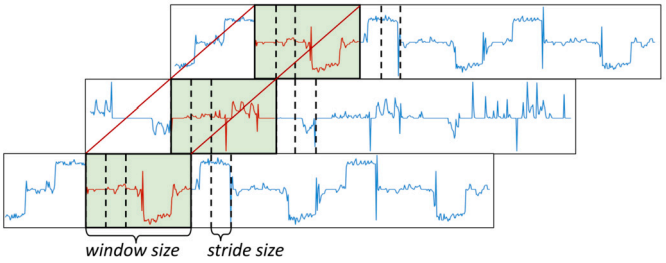


Fig. 3. Illustration of sequence segmentation for given hand-writing signal. Blue curves represent the independent dynamic features in the handwriting signal, green areas depict the segmented patch. The *window size* determines the patch length, and the *stride size* controls the extent of overlap between adjacent patches.

(w) and *stride size* (s). The window size (called temporal windows [31]) indicates the length of the cropped patches. Through the ablation experiment, a fixed window of 128 is employed in the experiments. The stride size controls the number of patch samplings, as well as the degree of overlap of the patches. A non-uniform stride size can also be used during the segmentation step. The process is described in Fig. 3. Moreover, such a segmentation scheme helps generate more data samples, which are essential for model training.

4.2. Network architecture

To establish the model architecture, we integrate multiple deep learning structures, including LSTM [32] and 1D CNNs. The detailed design of the LSTM-CNN is illustrated in Fig. 4, consisting of two essential components: the LSTM block and 1D CNN blocks.

4.2.1. LSTM block

The prominent LSTM [32] network has been extensively employed in processing diverse biomedical data, including electroencephalogram signals, electrocardiogram signals, genetic sequences, and other related domains [33,34]. Unlike conventional feed-forward neural networks, LSTM employs internal memory to process incoming inputs and effectively integrates longer temporal signals. Specifically, LSTM comprises a distinct set of memory units, where the current input data and prior states influence the output of the next state. This method enables the capture of temporal features from historical information in handwriting signals. In this study, the LSTM block comprises a single LSTM layer, composed of 128 memory units, where the input sequence size is 128×5. Each memory unit is equipped with standard cells that incorporate input gates, output gates, and forget gates. These gate mechanisms efficiently control information flow. With these capabilities, each cell can effectively preserve the desired values over extended time intervals. Furthermore, we add an efficient concatenation operation between the LSTM block and the subsequent 1D CNN blocks in "Stage 1". This operation concatenates the original input with the output of the LSTM block along the feature dimension, with an output size of 128×10, before inputting them into the 1D CNN block.

4.2.2. CNN block

The 1D CNN has recently emerged as a promising method, demonstrating state-of-the-art performance in biomedical data classification and early diagnosis [35]. Moreover, 1D and 2D CNNs share similar network structures, and the main distinction is that the convolutional operation of 1D CNNs is performed in only one direction. This fact implies that, given identical conditions (including configuration, network architecture, and hyperparameters), the computational complexity of a 1D CNN is significantly lower than that of its 2D counterpart. For example, convolving an input array with dimensions $M \times M$ by a $K \times K$ kernel has the $\sim\mathcal{O}(M^2K^2)$ computational complexity, whereas in cases of 1D convolution (with the identical dimensions, M and K), the complexity is only $\sim\mathcal{O}(MK)$. As depicted in Fig. 4, two convolutional blocks are employed in "Stage 2". Each block performs 1D convolution with the rectified linear unit (ReLU) activation, followed by a 1D max-pooling layer. Specifically, multiple filters of the same size are used in each convolutional block to capture information at various temporal scales. The initial convolutional block employs 16 filters with a kernel size of 3 and a stride of 2; subsequent convolutional blocks incorporate 32 filters, all with the same kernel size of 3 and a stride of 2. Finally, as depicted in Fig. 4(c), LSTM-CNN uses a fully connected block to discriminate handwriting signals. The dropout layer temporarily removes nodes from the network with a probability of 0.5 during the training phase.

4.3. Inference diagnosis

Finally, we employ a majority voting scheme for inference diagnosis. More specifically, we segment the acquired handwriting signal sample s into a set of fixed-length patches $\mathcal{P} = \{p_i, i = 1, 2, \ldots, n\}$ based on the mentioned data processing technique. Once these patches are obtained, the LSTM-CNN model predicates the corresponding classification result for each patch. Given a threshold $\alpha \in (0, 1)$, r_s represents the percentage of patches predicted as PD in the patches dataset. If $r_s \geq \alpha$, the handwriting signal s is predicated as PD, otherwise it is predicated as HC. Unless otherwise stated in the subsequent experiments, we set the threshold $\alpha = 0.5$ to indicate that the voting scheme should be consistent with the majority of reliable predictions. More details regarding choosing α are discussed in Section 5.3.

5. Experiments

This section presents the quantitative experiments to evaluate the performance of the proposed LSTM-CNN model. The five-fold cross-validation technique is deduced in the context of handwriting tasks. For the empirical analysis, we adopt three metrics: accuracy, sensitivity, specificity [36].

Implementation We implement LSTM-CNN using a Pytorch library [37]. LSTM-CNN is trained using an Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$, with a batch size of 16. The initial learning rate is 0.001, dropping by 10x in the 25, 50, 100 epochs. For binary classification (PD and HC), we use binary cross-entropy loss. All experiments are implemented on a desktop computer equipped with an Intel i7-11700 K 3.60 GHZ, 32 GB RAM and an 8 GB Nvidia RTX3070Ti GPU.

5.1. Comparison analysis

Table 1 compares the proposed model with methods widely adopted [8,11] for PD diagnosis on the DraWritePD dataset. All diagnostic methods are classified into model-based and deep learning-based methods, according to the differences in model architecture. We implement model-based methods in Python using the Scikit-learn library [38]. Additionally, the grid search algorithm is employed to optimize the hyperparameters. Specifically, (i) k-Nearest Neighbors (KNN): the possible number of neighbours is $K = [3, 5, 7, 9, 11]$. (ii) SVM: the radial basis function kernel is employed, the optimization range of kernel gamma γ and penalty parameter C are $\gamma = \left[2^{-2}, 2^{-1}, 2^{0}, 2^{1}, 2^{2}\right]$ and $C = \left[2^{-3}, 2^{-2}, 2^{-1}, 2^{0}, 2^{1}\right]$, respectively. (iii) Adaboost and RandomForest(RF): the possible number of decision trees N and the optimization range of maximum depth D are $N = [5, 10, 15, 20, 50]$ and $D = [2, 4, 8, 16, 32]$, respectively. On the other hand, we implement deep learning-based methods using the PyTorch framework [37]. Specifically, (i) CNN: the architecture is a standard and compact neural network, comprising just two 1D convolutional layers and fully connected layers. (ii) RNN-CNN and GRU-CNN: their architectures contain a layer of memory units stacked on top of the aforementioned CNN, and the only distinction from LSTM-CNN lies in the variation of memory units.

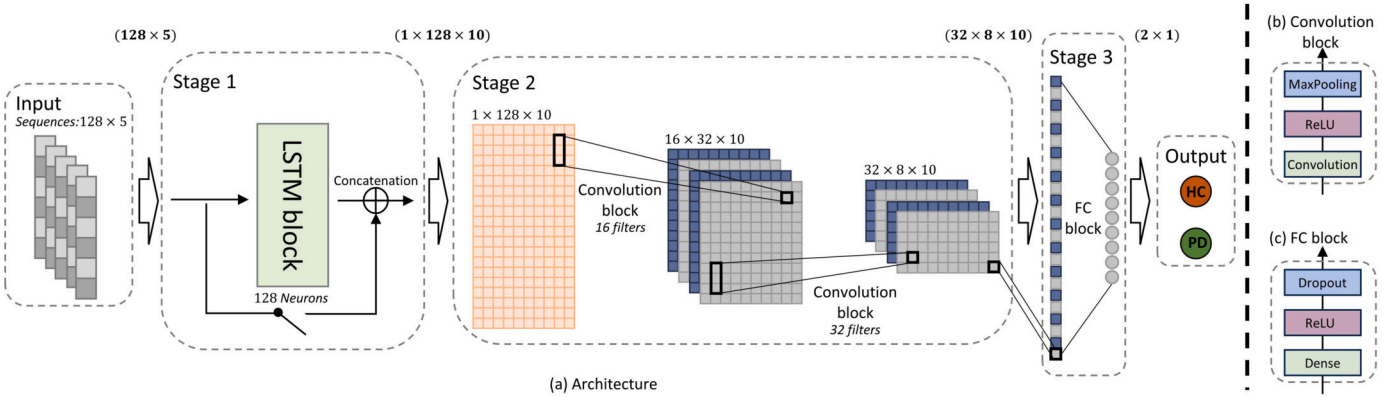


Fig. 4. Network architecture of LSTM-CNN for handwriting signal classification. The model input is a sequence (128×5) of length 128 and each time-step is a vector of five features. The model begins with a single LSTM block consisting of 128 recurrent units, followed by a concatenation operation that concatenates the input and LSTM output along the feature dimension. Subsequently, two standard one-dimensional convolutional blocks, one with 16 filters and another with 32 filters, both employing a stride of 2 and a kernel size of 3, are applied. Finally, the output layer comprises a single neuron with softmax activation to predict one of two classes (PD or HC).

Table 1
Quantitative comparison with various classification models on the DraWritePD dataset.

Method / Metric	Model-based methods				Deep learning-based methods			
	KNN	SVM	Adaboost	RF	CNN	RNN-CNN	GRU-CNN	LSTM-CNN(Ours)
Accuracy (%)	92.0 / 84.6	93.6 / 87.7	93.6 / 88.5	93.6 / 86.2	91.5 / 92.8	92.8 / 93.1	94.6 / 94.4	**96.2** / **95.2**
Sensitivity (%)	81.8 / 72.7	87.4 / 79.2	93.0 / 81.8	93.0 / 84.3	80.0 / 83.6	83.6 / 85.5	89.1 / 87.3	**94.5** / **89.1**
Specificity (%)	**100** / 93.3	96.8 / 94.3	94.3 / 93.3	94.3 / 88.7	**100** / **100**	**100** / 98.7	98.7 / **100**	97.3 / **100**
Params (K)	–	–	–	–	42.69	83.71	83.83	83.89
FLOPs (K)	–	–	–	–	202.37	446.21	542.21	590.21

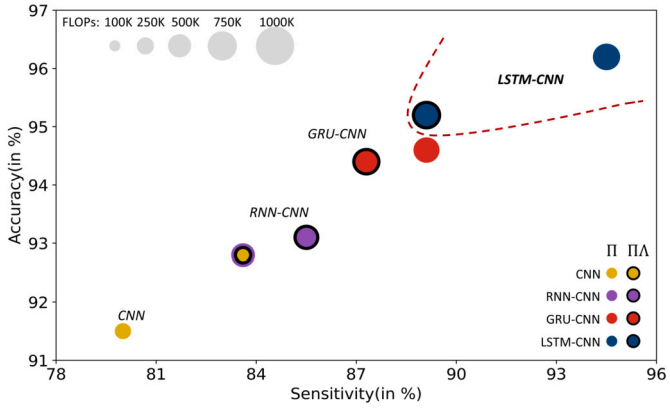


Fig. 5. Trade-off between performance and efficiency: proposed method vs. comparison methods.

Table 2
Time consumption of LSTM-CNN at each inference diagnosis stage (in seconds).

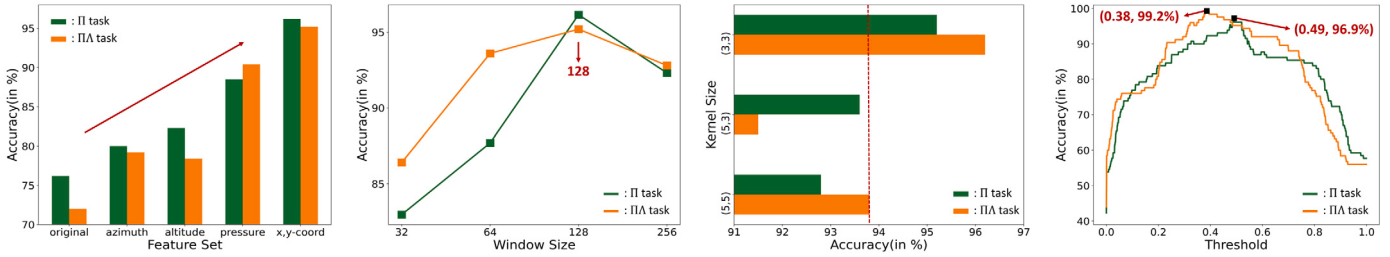
Case	Size	Loading	Processing	Diagnosing	Total
1	667	0.08491	0.00197	0.01097	0.10645
2	3128	0.09081	0.00299	0.01494	0.14963
3	7122	0.09108	0.00897	0.01795	0.16689
4	9471	0.09984	0.01097	0.02294	0.18761
5	12787	0.08614	0.01696	0.02690	0.19087
6	18618	0.09197	0.01995	0.03291	0.21960

For comparison, we report the results of model-based and other deep learning-based methods. The pairwise results are composed of a sequence of results in the task dataset Π and the task dataset ΠΛ. From this table we observe the following: (i) the proposed model significantly outperforms the existing model-based methods. This demonstrates the effectiveness of the proposed method in exploiting neural network architectures for the diagnosis of PD. (ii) The proposed model provides superior results among the deep learning-based methods. Furthermore, performance enhancements, especially in the Π task, are significant compared to other hybrid architectures incorporating memory units. This confirms the effectiveness of such a hybrid architecture, where memory units, especially LSTM, contribute to more discriminative temporal features. (iii) while other compact deep learning-based architectures are slightly smaller than the proposed architecture, they are outperformed by our model by a significant margin in terms of all performance metrics. (iv) significant performance disparities of model-based methods are observed in different drawing tasks. This reflects to some extent that the diagnosis of model-based methods is affected by hand-drawn patterns. The deep learning performance on these two tasks is basically the same, which indicates that the neural network may be free from the influence of the shape of external hand drawing patterns and is more inclined to extract more abstract intrinsic features. Note that the proposed model is lightweight enough to compete with other established ones such as AlexNet [26], ResNet [39], Transformer [40].

5.2. Complexity analysis

We evaluate the efficiency of the proposed LSTM-CNN from the perspectives of computational complexity and inference time. As presented in Table 1, the LSTM-CNN exhibits an impressively low parameter count of only 0.084M, accompanied by a mere 0.59 m floating point operations (FLOPs). This outcome theoretically indicates that the model possesses exceptional lightweight characteristics and boasts an exceptionally low computational complexity. Furthermore, we integrate diagnostic performance metrics with computational complexity to conduct a comprehensive evaluation. As illustrated in Fig. 5, the LSTM-CNN demonstrates significantly superior diagnostic performance while maintaining an acceptable level of computational complexity.

We also conduct on-site monitoring of the actual runtime for various stages involved in the inference diagnosis. Table 2 presents 6



(a) input features (b) temporal window (c) kernel size (d) threshold

Fig. 6. Quantitative comparison of various configurations: (a) evaluating the contribution of input feature sets, (b) comparing model performance under different window lengths, (c) choosing the appropriate convolution kernel size, and (d) illustrating model performance as a function of diagnostic threshold.

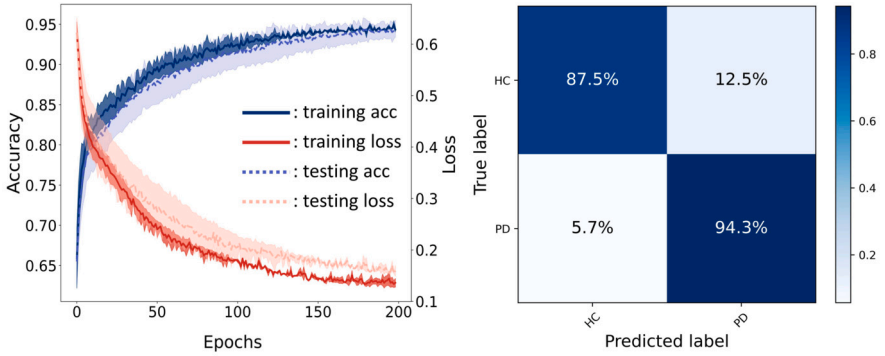


Fig. 7. Training process and Confusion matrix of LSTM-CNN on the PaHaW dataset.

Table 3
The influence of "concatenation" operation on model performance.

Metric \ Concatenation	✗	✓
Accuracy (%)	93.8 / 95.2	96.2 / 95.2
Sensitivity (%)	90.9 / 89.1	94.5 / 89.1
Specificity (%)	94.7 / 100	97.3 / 100

cases of inference diagnosis time, where each case displays the signal length, loading time for LSTM-CNN, pre-processing time, model diagnosis time, and overall duration. We observe that the overall diagnosis duration does not exceed 0.3 s, while the model prediction is nearly real-time, taking only 0.03 s. This observation further substantiates the lightweight and efficient nature of the proposed LSTM-CNN. Note that the algorithm framework is not specifically optimized in this study, and GPUs are not used in the inference diagnostic tests.

5.3. Ablation studies

Input feature set We conduct comparative experiments to evaluate the contribution of the handwriting feature on model performance. Specifically, we construct four distinct input feature sets by applying the forward difference operation to each individual feature while keeping the remaining features unchanged. It is important to emphasize that the features (x-coordinate and y-coordinate) collectively determine the coordinates of the data point, hence, they are considered as a whole. We include the original features as one additional input feature set, serving as the baseline. As illustrated in Fig. 6 (a), the raw feature set report the lowest classification rates. Not surprisingly, the geometric features (x-coordinate and y-coordinate), which contribute most to the aforementioned derived feature set, exhibit the top accuracy for all tasks [14–18].

Temporal window length Given that motor symptoms in patients with PD manifest primarily through local information obtained from the handwriting signal, careful selection of an appropriate window length (w) is imperative. In this study, we select four different window lengths for comparison, considering the distribution of signal lengths in the datasets. Fig. 6 (b) reveals that longer window lengths are more likely to yield superior diagnostic performance. Furthermore, the trend of increasing and then decreasing curves for accuracy is evident, with the optimal performance achieved at $w = 128$.

Model architecture Table 3 compares the effect of "concatenation" on the model architecture. The pairwise results in the table are sequentially composed for the task Π and the task ΠΛ. This table indicates that the incorporation of the "concatenation" operation maintains a stable model performance in the ΠΛ task, and yields an additional improvement of 2% to 4% in the model performance for the Π task. Moreover, we conducted comparative experiments on the combination of convolutional kernel sizes in LSTM-CNN. Fig. 6 (c) illustrates three combinations: 3 and 3, 5 and 3, and 5 and 5. The results show that in combination with 3 and 3, the model achieves optimal diagnostic performance on both handwriting tasks.

Threshold selection To explore the optimal performance of LSTM-CNN, we conduct additional supplementary experiments on threshold α in the majority voting scheme. Fig. 6 (d) shows that with a threshold of 0.49, the highest classification accuracy achieved on Π task is 96.9%; with a threshold of 0.38, the highest classification accuracy achieved on ΠΛ task is 99.2%.

5.4. Robustness validation

To validate the robustness of the proposed method, we performed additional experiments using the spiral handwriting task on the public PaHaW dataset [15,17]. This data set is not employed in the configuration of the proposed system; therefore, we adopt the optimal configuration described in Section 5.3. Fig. 7 illustrates the model's training process and confusion matrix. Furthermore, in Table 4, we analyse the state-of-the-art studies that use this specific data set to con-

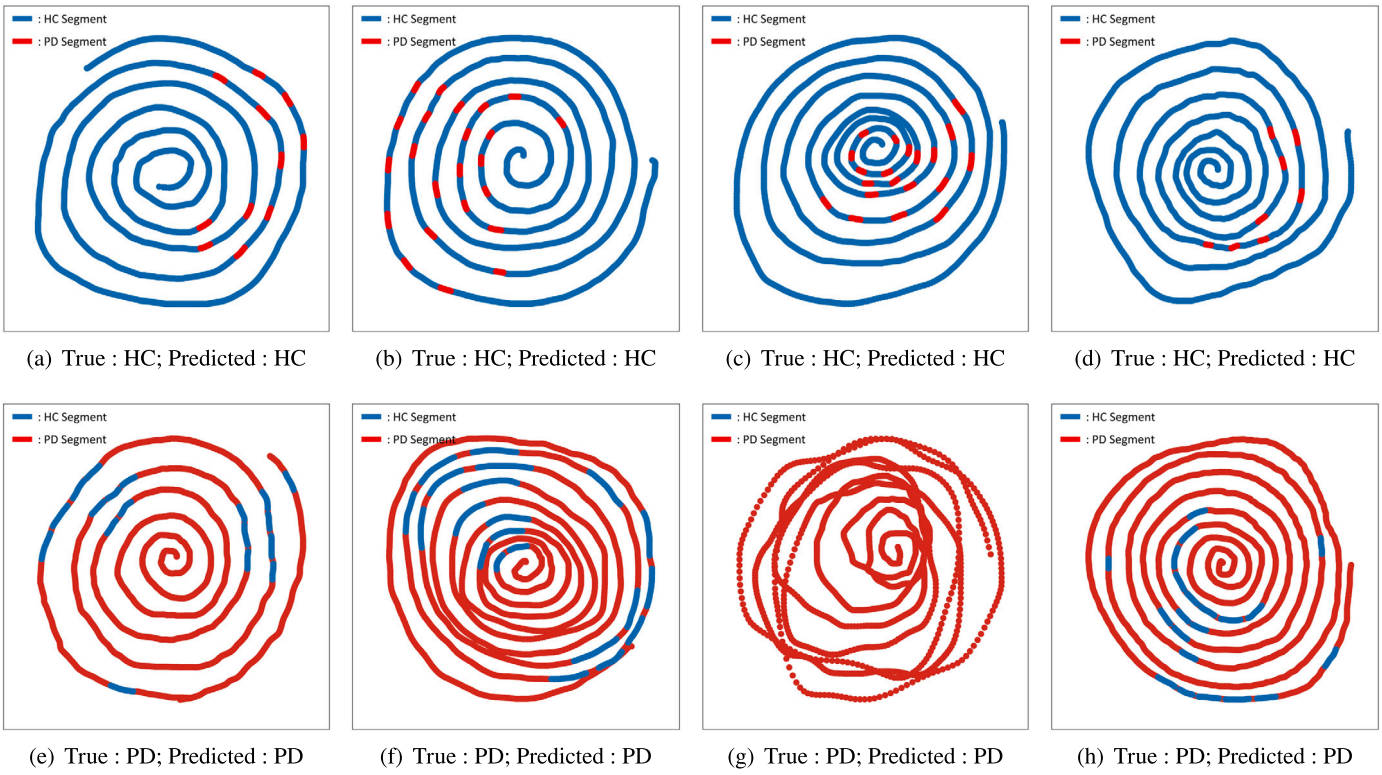


Fig. 8. Visualization diagnostic cases from the PaHaW dataset. (a)-(d): the cases are correctly diagnosed as HC, (e)-(h): the cases are correctly diagnosed as PD. Blue segments represent model predictions as HC, red segments represent model predictions as PD.

Table 4
Quantitative comparison with state-of-the-art works on the PaHaW dataset.

Method	Features	Models	Accuracy (%)	Year
Drotár et al. [17]	kinematic, spatio-temporal and pressure features	RF, SVM	62.8	2016
Angelillo et al. [41]	velocity-based features	SVM	53.8	2019
Impedovo [25]	velocity-based features	SVM	97.33	2019
Diaz et al. [21]	static images with dynamically enhanced	2D CNN+SVM	75.0	2019
Valla et al. [18]	derivative-based, angle-type, and integral-like features	KNN, SVM	84.9	2022
Ours	dynamic handwriting patterns	LSTM-CNN	**90.7**	–

textualise the results. Furthermore, since different classifiers (support vector machines, random forests, etc.) are used in these studies, reporting different classification accuracies, for a fair comparison, we report here only the best results of these studies. As is evident from this table, the proposed method exhibits superior performance compared to previously proposed methods that rely on traditional handcrafted features or 2D image recognition. However, it is worth pointing out that Impedovo [25] used the non-nested validation scheme. This means that while remarkable, the contribution provided still suffers from the feature selection bias. This result further confirms the effectiveness of the proposed method as a candidate solution for practical use in clinical settings. In addition, in this work, we primarily use this data set to validate the robustness of the proposed method.

6. Discussion

In this study, we validate the potential of the analysis of handwriting signal segments to help with the early diagnosis of PD and propose an efficient diagnostic method accordingly. Furthermore, we visualise diagnostic results, enabling intuitive interpretation in clinical applications without the need for specific computational expertise. Compared to the "holistic" analysis of subtle variations in the signal [17,19], methods based on handwriting signal segments can better quantify local motor symptoms, such as bending, horizontal, and certain angles of spirals during hand drawing. And motor symptoms at these critical moments may be crucial for the diagnosis of PD [7]. Furthermore, significant performance disparities of model-based methods are observed in different drawing tasks in Table 1, indicating their susceptibility to external shapes of handwriting tasks. On the contrary, the proposed LSTM-CNN model tends to extract generalisable features from different handwriting signals. Second, based on ablation experiments presented in Fig. 6(a), it is observed that geometric features (x, y coordinates) contain more discriminative temporal information compared to kinematic features, followed by the second most common pressure features. These findings are consistent with previous studies [14–18]. Perhaps due to the fact that patients are likely to introduce more positional changes, making their drawings much more nonfluent [8]. Finally, we provide an intuitive diagnosis on spiral drawing tasks. It can be seen from Fig. 8 that most abnormal diagnoses are concentrated within certain angle ranges of the spiral, indicating that hand drawing in these angle ranges may contain more distinctive information. Due to the highly individualised nature of PD, we are currently unable to provide universally applicable explanations. However, we speculate that this may be related to PD micrographia [42], which leads to inadequate coordination of the fingers, wrists, and arms at these specific angles [43]. In the future, investigating the correlation between abnormal segments and patient information (e.g., UPDRS V [44], LED [45]) may potentially reveal novel findings.

However, the proposed method can help healthcare professionals focus their attention to these critical moments to enhance diagnostic efficacy, when combined with domain expertise. Furthermore, the lightweight, fast and efficient attributes of our method offer opportunities for practical applications in the field of medicine.

7. Conclusions

In this article, we propose an efficient method to analyse handwriting sequence segments to facilitate the early diagnosis of PD. Furthermore, a hybrid neural network is designed that incorporates LSTM and one-dimensional convolution, demonstrating outstanding performance while maintaining a lightweight structure. The systemic experimental results substantiate that the proposed method offers efficient diagnostic performance, minimal computational requirements, and strong robustness compared to current state-of-the-art methods. Finally, we visualise the diagnostic results, improving the credibility of the model diagnosis and promoting practical applications in the medical field. This study has several limitations and suggestions for further improvement. The development of large benchmark datasets remains one of the ongoing issues for the community in this field. This applies not only to PD, but also to other neurodegenerative diseases. Some studies have attempted to address this issue through data augmentation techniques [46] or transfer learning [47]. In the future, our work strives to overcome this obstacle by providing additional databases and new methods for data augmentation. Furthermore, correlation analysis can be employed to further explore the interpretability of handwriting segments, facilitating their application in practical medical settings.

CRediT authorship contribution statement

Xuechao Wang: Writing – original draft, Validation, Methodology, Formal analysis. **Junqing Huang:** Writing – original draft, Validation, Methodology. **Marianna Chatzakou:** Writing – review & editing, Methodology, Investigation. **Kadri Medijainen:** Data curation. **Aaro Toomela:** Data curation. **Sven Nõmm:** Writing – review & editing, Supervision, Methodology, Investigation, Data curation. **Michael Ruzhansky:** Writing – review & editing, Supervision, Methodology.

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.

Acknowledgements

This study was supported by the grant PRG 957 of the Estonian Research Council. This work was supported by the Estonian Research Council grant PRG 2100. It was also partially supported by the FWO Odysseus 1 grant G.0H94.18N: Analysis and Partial Differential Equations, and the Methusalem programme of the Ghent University Special Research Fund (BOF) (Grant number 01M01021). Michael Ruzhansky is also supported by EPSRC grant EP/R003025/2. M. Chatzakou is a postdoctoral fellow of the Research Foundation – Flanders (FWO) under the postdoctoral grant No 12B1223N.

References

[1] Oleh Hornykiewicz, Biochemical aspects of Parkinson's disease, Neurology 51 (2 Suppl 2) (1998) S2–S9.
[2] Daniel L. Murman, Early treatment of Parkinson's disease: opportunities for managed care, Amer. J. Manag. Care 18 (7) (2012) S183.
[3] Robert A. Hauser, Early pharmacologic treatment in Parkinson's disease, Amer. J. Manag. Care 16 (2010), S100-7.
[4] Sigurlaug Sveinbjornsdottir, The clinical symptoms of Parkinson's disease, J. Neurochem. 139 (2016) 318–324.
[5] Andrew J. Hughes, Susan E. Daniel, Siobhan Blankson, Andrew J. Lees, A clinicopathologic study of 100 cases of Parkinson's disease, Arch. Neurol. 50 (2) (1993) 140–148.
[6] E. Ray Dorsey, Alexis Elbaz, Emma Nichols, Nooshin Abbasi, Foad Abd-Allah, Ahmed Abdelalim, Jose C. Adsuar, Mustafa Geleto Ansha, Carol Brayne, Jee-Young J. Choi, et al., Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016, Lancet Neurol. 17 (11) (2018) 939–953.
[7] Eli Carmeli, Hagar Patish, Raymond Coleman, The aging hand, J. Gerontol., Ser. A, Biol. Sci. Med. Sci. 58 (2) (2003) M146–M152.
[8] Mathew Thomas, Abhishek Lenka, Pramod Kumar Pal, Handwriting analysis in Parkinson's disease: current status and future directions, Mov. Disord. Clin. Pract. 4 (6) (2017) 806–818.
[9] Ibtissame Aouraghe, Ghizlane Khaissidi, Mostafa Mrabti, A literature review of online handwriting analysis to detect Parkinson's disease at an early stage, Multimed. Tools Appl. (2022) 1–26.
[10] Aa Schrag, Y.B. Ben-Shlomo, Niall Quinn, How valid is the clinical diagnosis of Parkinson's disease in the community?, J. Neurol. Neurosurg. Psychiatry 73 (5) (2002) 529–534.
[11] Ibtissame Aouraghe, Ghizlane Khaissidi, Mostafa Mrabti, A literature review of online handwriting analysis to detect Parkinson's disease at an early stage, Multimed. Tools Appl. 82 (8) (2023) 11923–11948.
[12] Muhammed Isenkul, Betul Sakar, Olcay Kursun, et al., Improved spiral test using digitized graphics tablet for monitoring Parkinson's disease, in: The 2nd International Conference on e-Health and Telemedicine (ICEHTM-2014), vol. 5, 2014, pp. 171–175.
[13] Jens Barth, Michael Sünkel, Katharina Bergner, Gerald Schickhuber, Jürgen Winkler, Jochen Klucken, Björn Eskofier, Combined analysis of sensor data from hand and gait motor function improves automatic recognition of Parkinson's disease, in: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2012, pp. 5122–5125.
[14] Peter Drotár, Jiří Mekyska, Irena Rektorová, Lucia Masarová, Zdeněk Smékal, Marcos Faundez-Zanuy, A new modality for quantitative evaluation of Parkinson's disease: in-air movement, in: 13th IEEE International Conference on Bioinformatics and Bioengineering, IEEE, 2013, pp. 1–4.
[15] Peter Drotár, Jiří Mekyska, Irena Rektorová, Lucia Masarová, Zdenek Smékal, Marcos Faundez-Zanuy, Analysis of in-air movement in handwriting: a novel marker for Parkinson's disease, Comput. Methods Programs Biomed. 117 (3) (2014) 405–411.
[16] Peter Drotár, Jiří Mekyska, Irena Rektorová, Lucia Masarová, Zdeněk Smékal, Marcos Faundez-Zanuy, Decision support framework for Parkinson's disease based on novel handwriting markers, IEEE Trans. Neural Syst. Rehabil. Eng. 23 (3) (2014) 508–516.
[17] Peter Drotár, Jiří Mekyska, Irena Rektorová, Lucia Masarová, Zdeněk Smékal, Marcos Faundez-Zanuy, Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease, Artif. Intell. Med. 67 (2016) 39–46.
[18] Elli Valla, Sven Nomm, Kadri Medijainen, Pille Taba, Aaro Toomela, Tremor-related feature engineering for machine learning based Parkinson's disease diagnostics, Biomed. Signal Process. Control 75 (2022) 103551.
[19] Clayton R. Pereira, Silke A.T. Weber, Christian Hook, Gustavo H. Rosa, Joao P. Papa, Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics, in: 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), IEEE, 2016, pp. 340–346.
[20] Clayton R. Pereira, Danilo R. Pereira, Gustavo H. Rosa, Victor H.C. Albuquerque, Silke A.T. Weber, Christian Hook, João P. Papa, Handwritten dynamics assessment through convolutional neural networks: an application to Parkinson's disease identification, Artif. Intell. Med. 87 (2018) 67–77.
[21] Moises Diaz, Miguel Angel Ferrer, Donato Impedovo, Giuseppe Pirlo, Gennaro Vessio, Dynamically enhanced static handwriting representation for Parkinson's disease detection, Pattern Recognit. Lett. 128 (2019) 204–210.
[22] Luis C.S. Afonso, Gustavo H. Rosa, Clayton R. Pereira, Silke A.T. Weber, Christian Hook, Victor Hugo C. Albuquerque, João P. Papa, A recurrence plot-based approach for Parkinson's disease identification, Future Gener. Comput. Syst. 94 (2019) 282–292.
[23] Sven Nõmm, Sergei Zarembo, Kadri Medijainen, Pille Taba, Aaro Toomela, Deep CNN based classification of the Archimedes spiral drawing tests to support diagnostics of the Parkinson's disease, IFAC-PapersOnLine 53 (5) (2020) 260–264.
[24] Jan Mucha, Jiri Mekyska, Zoltan Galaz, Marcos Faundez-Zanuy, Karmele Lopez-de Ipina, Vojtech Zvoncak, Tomas Kiska, Zdenek Smekal, Lubos Brabenec, Irena Rektorova, Identification and monitoring of Parkinson's disease dysgraphia based on fractional-order derivatives of online handwriting, Appl. Sci. 8 (12) (2018) 2566.
[25] Donato Impedovo, Velocity-based signal features for the assessment of Parkinsonian handwriting, IEEE Signal Process. Lett. 26 (4) (2019) 632–636.
[26] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012).
[27] Hamzah Luqman, Elalfy Elsayed, Utilizing motion and spatial features for sign language gesture recognition using cascaded CNN and LSTM models, Turk. J. Electr. Eng. Comput. Sci. 30 (7) (2022) 2508–2525.
[28] Luiz C.F. Ribeiro, Luis C.S. Afonso, Joao P. Papa, Bag of samplings for computer-assisted Parkinson's disease diagnosis based on recurrent neural networks, Comput. Biol. Med. 115 (2019) 103477.

[29] Moises Diaz, Momina Moetesum, Imran Siddiqi, Gennaro Vessio, Sequence-based dynamic handwriting analysis for Parkinson's disease detection with one-dimensional convolutions and BiGRUs, Expert Syst. Appl. 168 (2021) 114405.
[30] Umut Özkaya, Şaban Öztürk, Farid Melgani, Levent Seyfi, Residual CNN+ Bi-LSTM model to analyze GPR B scan images, Autom. Constr. 123 (2021) 103525.
[31] Artur Jordao, Antonio C. Nazare Jr, Jessica Sena, William Robson Schwartz, Human activity recognition based on wearable sensor data: a standardization of the state-of-the-art, arXiv preprint, arXiv:1806.05226, 2018.
[32] Alex Sherstinsky, Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network, Phys. D, Nonlinear Phenom. 404 (2020) 132306.
[33] Shuting Xu, Oliver Faust, Silvia Seoni, Subrata Chakraborty, Prabal Datta Barua, Hui Wen Loh, Heather Elphick, Filippo Molinari, U. Rajendra Acharya, A review of automated sleep disorder detection, Comput. Biol. Med. 150 (2022) 106100.
[34] Süleyman Akdağ, Fatma Kuncan, Yilmaz Kaya, A new approach for congestive heart failure and arrhythmia classification using downsampling local binary patterns with LSTM, Turk. J. Electr. Eng. Comput. Sci. 30 (6) (2022) 2145–2164.
[35] Serkan Kiranyaz, Onur Avci, Osama Abdeljaber, Turker Ince, Moncef Gabbouj, Daniel J. Inman, 1D convolutional neural networks and applications: a survey, Mech. Syst. Signal Process. 151 (2021) 107398.
[36] Margherita Grandini, Enrico Bagli, Giorgio Visani, Metrics for multi-class classification: an overview, arXiv preprint, arXiv:2008.05756, 2020.
[37] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al., Pytorch: an imperative style, high-performance deep learning library, Adv. Neural Inf. Process. Syst. 32 (2019).
[38] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al., Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[39] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[40] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).
[41] Maria Teresa Angelillo, Donato Impedovo, Giuseppe Pirlo, Gennaro Vessio, Performance-driven handwriting task selection for Parkinson's disease classification, in: AI* IA 2019–Advances in Artificial Intelligence: XVIIIth International Conference of the Italian Association for Artificial Intelligence, Rende, Italy, November 19–22, 2019, Proceedings 18, Springer, 2019, pp. 281–293.
[42] J.E. McLennan, K. Nakano, H.R. Tyler, R.S. Schwab, Micrographia in Parkinson's disease, J. Neurol. Sci. 15 (2) (1972) 141–152.
[43] Diptaman Chatterjee, Jeffrey H. Kordower, Immunotherapy in Parkinson's disease: current status and future directions, Neurobiol. Dis. 132 (2019) 104587.
[44] S.R.L.E. Fahn, Unified Parkinson's disease rating scale, in: Recent Developments in Parkinson's Disease, 1987, pp. 153–163.
[45] Jee-Young Lee, Jae Woo Kim, Won Yong Lee, Jong-Min Kim, Tae-Beom Ahn, Han-Joon Kim, Jinwhan Cho, Beom S. Jeon, Daily dose of dopaminergic medications in Parkinson disease: clinical correlates and a posteriori equation, Neurol. Asia 15 (2010).
[46] Erik Dzotsenidze, Elli Valla, Sven Nõmm, Kadri Medijainen, Pille Taba, Aaro Toomela, Generative adversarial networks as a data augmentation tool for CNN-based Parkinson's disease diagnostics, IFAC-PapersOnLine 55 (29) (2022) 108–113.
[47] Matej Gazda, Máté Hireš, Peter Drotár, Multiple-fine-tuned convolutional neural networks for Parkinson's disease diagnosis from offline handwriting, IEEE Trans. Syst. Man Cybern. Syst. 52 (1) (2021) 78–89.