Journal Pre-proof

Sequence-based Dynamic handwriting analysis for Parkinson's disease detection with one-dimensional convolutions and BiGRUs

Moises Diaz, Momina Moetesum, Imran Siddiqi, Gennaro Vessio

PII: S0957-4174(20)31075-7
DOI: https://doi.org/10.1016/j.eswa.2020.114405
Reference: ESWA 114405

To appear in: *Expert Systems With Applications*

Received date : 4 August 2020
Revised date : 5 November 2020
Accepted date : 27 November 2020

Please cite this article as: M. Diaz, M. Moetesum, I. Siddiqi et al., Sequence-based Dynamic handwriting analysis for Parkinson's disease detection with one-dimensional convolutions and BiGRUs. *Expert Systems With Applications* (2020), doi: https://doi.org/10.1016/j.eswa.2020.114405.

Sequence-based Dynamic Handwriting Analysis for Parkinson's Disease Detection with One-dimensional Convolutions and BiGRUs

Moises Diaz[a], Momina Moetesum[b], Imran Siddiqi[b], Gennaro Vessio[c,*]

[a] *Universidad del Atlántico Medio, Las Palmas de Gran Canaria, Spain*
[b] *Vision & Learning Lab, Bahria University Islamabad, Pakistan*
[c] *Department of Computer Science, University of Bari "Aldo Moro", Italy*

Abstract

Parkinson's disease (PD) is commonly characterized by several motor symptoms, such as bradykinesia, akinesia, rigidity, and tremor. The analysis of patients' fine motor control, particularly handwriting, is a powerful tool to support PD assessment. Over the years, various dynamic attributes of handwriting, such as pen pressure, stroke speed, in-air time, etc., which can be captured with the help of online handwriting acquisition tools, have been evaluated for the identification of PD. Motion events, and their associated spatio-temporal properties captured in online handwriting, enable effective classification of PD patients through the identification of unique sequential patterns. This paper proposes a novel classification model based on one-dimensional convolutions and Bidirectional Gated Recurrent Units (BiGRUs) to assess the potential of sequential information of handwriting in identifying Parkinsonian symptoms. One-dimensional convolutions are applied to raw sequences as well as derived features; the resulting sequences are then fed to BiGRU layers to achieve the final classification. The proposed method outperformed state-of-the-art approaches on the PaHaW dataset and achieved competitive results on the NewHandPD dataset.

Keywords: Parkinson's disease, Dynamic handwriting analysis, Recurrent

⋆This document is a collaborative effort.

*Corresponding author
Email addresses: moises.diaz@atlanticomedio.es (Moises Diaz), reach.momina@gmail.com (Momina Moetesum), imran.siddiqi@bahria.edu.pk (Imran Siddiqi), gennaro.vessio@uniba.it (Gennaro Vessio)

neural networks, Computer-aided diagnosis

1. Introduction

Parkinson's disease (PD) is one of the most widespread and most disabling neurodegenerative disorders; it adversely affects the structure and functions of brain areas resulting in a gradual cognitive, behavioral, and functional decline (Ascherio & Schwarzschild 2016). At present, there is no cure, and the progressive deterioration of the patient can only be somehow managed during disease progression. Nevertheless, early diagnosis of PD could be crucial from the perspective of proper medical treatment to be administered as well as to evaluate the effectiveness of new drug treatments at prodromal stages (Bhat et al. 2018). Moreover, the assessment of signs and manifestations of this specific disease is useful for its diagnostic differentiation from similar disorders, and for monitoring and tracking its progression as the disease advances. With this aim, over the years, a growing interest of the community has been observed in computer-aided diagnosis (Parisi et al. 2018, Ali et al. 2019). Such intelligent systems can effectively assist clinicians at the point of care, providing novel decision support tools, while reducing expenditure on public health.

Alterations in the brain caused by PD, such as neuronal loss, synaptic dysfunction, brain atrophy, etc., among others, can result in a malfunction of the motor system and its components. This is particularly manifested in performance impairment of previously learned motor skills. In this view, a unique role in the context of PD assessment can be safely assumed for handwriting. Handwriting is a complex activity involving perceptual-motor as well as cognitive components, the changes of which can be considered a promising *biomarker* for disease assessment (De Stefano et al. 2019, Vessio 2019, Faundez-Zanuy et al. 2020). Indeed, there is a growing body of knowledge which provides evidence that the automatic discrimination between unhealthy and healthy individuals can be accomplished through the use of simple and easy-to-perform handwriting tasks, e.g. (Rosenblum et al. 2013, Drotár et al. 2016, Ammour et al. 2020). Developing a handwriting-based decision support tool is desirable, as it can provide a non-invasive, real-time, and low-cost solution to support the standard clinical evaluations carried out by human experts.

Within this research direction, on-line (*dynamic*) systems based on the use of a digitizing tablet can be adopted. Such a device allows one to capture not only temporal and spatial variables of handwriting but also the pressure exerted by the pen over the writing surface, as well as measures of pen orientation and inclination. Moreover, this technology can acquire pen movement not only while the pen is in contact with the writing surface, but also when the pen is in the proximity of the surface, i.e. "in-air". Contrary to off-line (*static*) features of handwriting, which can be analyzed after the writing process has already occurred, dynamic handwriting analysis deals with those features that can be acquired during the execution of the writing process. This can provide the system with rich dynamic information that can be exploited for disease diagnosis (Drotár et al. 2014*a*).

When designing such a system, a crucial step involves choosing the most appropriate features to describe handwriting. By directly feeding a classic statistical learning classifier with the time-series raw data, as acquired by the tablet, the model would suffer the burden of high dimensionality and thus overfitting. For this reason, several dynamic features have been derived from data in their raw form, ranging from traditional kinematic and spatio-temporal variables of handwriting to less common measures based, for example, on entropy and signal-to-noise ratio, e.g. (Rosenblum et al. 2013, Drotár et al. 2014*b*, Impedovo 2019). It is worth noting that, to obtain complete statistical representations of the available features, mathematical functions of the feature vector (including mean, median, standard deviation, and so on) are generally computed. However, although this "holistic" approach can help the model find effective decision frontiers in the feature space, on the other hand, it may lose relevant information, as an arbitrarily long sequence is condensed into single-valued features.

Another approach to representing handwritten patterns is to use features automatically learned by deep learning models. Some recent works based on Convolutional Neural Networks (CNNs) address the automatic extraction of features from two-dimensional static images by exploiting dynamic information of the handwriting (Pereira et al. 2018, Diaz et al. 2019*a*). While this approach represents a robust alternative to manually engineered features, it also provides only a holistic view of the handwritten patterns under study; moreover, since it is black-box, this approach obfuscates the meaning of the features employed and their correlation with the concomitant disease.

An alternative way to process the time-series data without losing relevant information, which has not yet been explored to its full extent in this domain, is to apply the sequence-based neural learning paradigm using Recurrent Neural Network (RNN) models. The on-line recordings captured during writing can exhibit unique time-dependent patterns, which can be exploited to discriminate PD patients from healthy controls. Instead of compressing the original data into single-valued features, as done by many researchers, we want to exploit the sequential nature of the data to explicitly take time into account and gain new insights into the dynamic handwriting/drawing process. Although traditional methods have proved useful without explicitly modeling time, RNNs are powerful tools for modeling data with temporal or sequential structures of variable length (Lipton et al. 2015). Two commonly used recurrent units include the Long-Short Term Memory (LSTM) (Hochreiter & Schmidhuber 1997) and Gated Recurrent Units (GRUs) (Cho et al. 2014). In recent years, systems based on these architectures have shown ground-breaking performance in traditionally challenging-to solve tasks, such as image captioning, language translation, and handwriting recognition, e.g. (You et al. 2016, Zhang et al. 2017).

The research presented in this paper represents a contribution to the state-of-the-art on sequence-based dynamic handwriting analysis for PD identification, extending our pilot study in this direction (Moetesum et al. 2020). More specifically, we apply one-dimensional convolution to the raw sequences (as well as derived features), to take advantage of the abundant temporal information from the handwriting samples. This not only results in a robust feature representation, but also serves to sub-sample these sequences to mitigate overfitting while reducing training time. The resulting feature sequences are then fed to Bidirectional GRU (BiGRU) layers to achieve the final classification. This approach is best suited for capturing the temporal sequence of the handwritten patterns, in which muscle contractions and irregular movements due to Parkinsonism may be reflected. In fact, a significant improvement in the identification rate compared to the state-of-the-art is observed, in a fair experimental comparison carried out on the same dataset, namely PaHaW (Drotár et al. 2016), on a task-by-task basis. Moreover, an analysis of the significance of features is also carried out as a function of the classification rates reported. Finally, it is worth noting that PaHaW is based on samples acquired through a digitizing tablet. To further evaluate

the robustness of our method, it was also tested on a dataset acquired via a smart pen, namely NewHandPD (Pereira, Weber, Hook, Rosa & Papa 2016).

The rest of this paper is organized as follows. Section 2 reviews the notable works related to this problem. Section 3 and 4 respectively describe the materials and methods used in this research. Section 5 reports and discusses the experimental results to highlight the effectiveness of the proposed method, while Section 6 draws conclusions and presents the final remarks.

2. Related Work

In the context of Parkinson's disease assessment, dynamic handwriting analysis has been applied to investigate several issues and has attracted growing interest from diverse research areas (psychology, neuroscience, computer science, and so on). A large part of the literature on this topic investigated fine motor control impairments. The analysis of changes in handwriting facilitated the understanding of the brain-body functional relationships and led to some recognizable patterns of the sensorimotor dysfunction associated with PD, e.g. (Teulings & Stelmach 1991, Smits et al. 2014, Senatore & Marcelli 2019). Many other works, e.g. (Eichhorn et al. 1996, Randhawa et al. 2013, Danna et al. 2019) have focused on studying the effects of medication on handwriting; analyzing the evolving patterns in handwriting can provide a useful tool for monitoring and tracking disease progression. More recently, significant research endeavors have been made towards the development of decision support tools to automatically discriminate between PD patients and healthy individuals (Rosenblum et al. 2013, Drotár et al. 2014*a*, Zham et al. 2017). This research has been particularly stimulated by the recent advances in machine (deep) learning techniques. The ultimate goal is to provide clinicians with a complementary approach to their standard evaluation, which is fast, non-invasive, and low-cost. The present work is part of this research direction.

Among the notable contributions to PD identification from computerized handwriting analysis, the most significant series of works has been reported by Drotár *et al.* All of their studies were carried out on the same dataset, i.e. PaHaW (Drotár et al. 2014*a*), which was subsequently made available to the community. In (Drotár et al. 2014*a*), the authors investigated the extent to which classification performance can be improved, considering not

only on-surface but also in-air movements as the two handwriting modalities appear to carry non-redundant information. In addition to computing conventional kinematic handwriting measures, such as velocity, acceleration, and jerk, Drotár et al. (2014*b*) also used relevant quantifiers based on entropy, signal energy, and empirical mode decomposition. These features provided novel insight and better understanding of the data. Subsequently, in (Drotár et al. 2016), the authors introduced additional fundamental features based on the pressure exerted over the writing surface. Specifically, they used the pressure values acquired by the tablet along with the rate at which the pressure signal changes over time.

The main factor contributing to the popularity of the PaHaW dataset in the research community is the collection of multiple handwriting tasks, ranging from the well-known Archimedes spiral drawing to word and sentence writing. Unfortunately, there are currently very few datasets freely available for research that provide multiple tasks, of varying degrees of complexity, performed by the same subjects. To better place our research in the literature panorama, we preferred to use this dataset. It is worth noting that, in all of the studies carried out by Drotár *et al.*, the spiral task was undertaken without any significant impact on classification. This may have been due to the use of measures suitable only for handwriting; on the contrary, visual features, such as those extracted by Convolutional Neural Network models (Diaz et al. 2019*a*, Moetesum et al. 2019), seem to overcome this issue.

Impedovo (2019) improved the results obtained on the PaHaW dataset by combining classic features with new velocity-based features. The extended feature set includes parameters obtained from the Sigma-Lognormal model (Ferrer et al. 2020), the Maxwell-Boltzmann distribution, and the Discrete Fourier Transform applied to the velocity profile of handwriting. Rios-Urrego et al. (2019), in addition to kinematic features, proposed to use geometrical and non-linear dynamic features. These features were proposed in the assumption that they are able to capture the irregularities of handwriting, which increase as the disease advances. In all the works discussed above, statistics computed on traditional hand-crafted dynamic features have been used to characterize PD.

Among other well-known contributions, Pereira, Weber, Hook, Rosa &

Papa (2016) introduced NewHandPD, a dataset of signals extracted from an electronic smart pen, which includes spiral and meander drawings. Each sensor of the pen outputs the overall signal acquired during the handwriting task, which can subsequently be represented as a time-series. The authors proposed to cast the problem of distinguishing PD from controls as an image recognition task through CNNs. Their strategy was to transform the signals provided by the smart pen into images. This research was one of the first applications of a deep learning-oriented approach to aid in the diagnosis of PD. The work was later extended in (Pereira et al. 2018) and (Afonso et al. 2019). In (Pereira et al. 2018), CNNs were employed to learn texture-oriented features directly from the time-series-based images. The central hypothesis was that these features could encode hand tremors during handwriting. In (Afonso et al. 2019), on the other hand, the recurrence plot technique was used to map the pen signals into the image domain; these images were then fed into a CNN to learn effective features. A recurrence plot enables the visualization of repeated events of higher dimensions through projections onto low-dimensional representations and can be exploited to identify PD subjects.

Recently, in (Diaz et al. 2019*a*), we proposed a "dynamically enhanced" representation of handwriting that consists of synthetically generated images obtained by jointly exploiting static and dynamic properties of handwriting. Specifically, we studied a static representation that embeds dynamic information based on drawing the points of the samples, instead of linking them, to preserve some velocity information, and adding pen-ups in the same way. The new handwriting representation, which was fed into CNNs to extract features automatically, was able to outperform the results obtained using static and dynamic handwriting separately on PaHaW. Unfortunately, although augmented with velocity and in-air information, the enhanced representation is still "static" and does not help the model reconstruct the temporal sequence of the handwriting movement.

More recently, Ribeiro et al. (2019) focused on the analysis of tremor, one of the most distinctive characteristics of PD. The authors proposed to learn temporal information from time-dependent signals by exploiting an RNN-based model along with an attention mechanism. The authors observed performance degradation due to long sequences, and the problem was addressed using a bag-of-sampling technique as a compact signal representation. Experimental results on the NewHandPD dataset compared favorably

with the previous literature. The study advocated the potential of sequential data analysis for PD identification and motivated us to further explore this research direction.

3. Materials

Two datasets have been considered in this work. Both are publicly available for research and include time-based sequences of people with Parkinson's disease. Moreover, they both contain a similar number of specimens, which makes experimentation more balanced. First, the PaHaW dataset (Drotár et al. 2014*a*) was used to adjust our system and find the best configuration. Second, the NewHandPD dataset (Pereira, Weber, Hook, Rosa & Papa 2016) was used as an additional test bed for our method, as it contains handwriting samples acquired not through a tablet but via a smart pen. These data, which have not been seen by our system, would then lead to confirm the robustness of our system.

3.1. PaHaW

The "Parkinson's disease handwriting database" (PaHaW) collects handwriting data of 37 PD patients and 38 age and gender-matched healthy control (HC) subjects (Drotár et al. 2014*a*). Participants were enrolled at the First Department of Neurology, Masaryk University, and the St. Anne's University Hospital, Brno, Czech Republic. All participants were right-handed, had completed at least ten years of education, and reported Czech as their native language. No significant between-group difference regarding age or gender was found. None of the subjects had a history or presence of any psychiatric symptom or disease affecting the central nervous system, with the exception of Parkinsonism in the PD group. Patients were only examined in their ON-state while taking dopaminergic medication, and, prior to acquisition, they were evaluated by a qualified neurologist. Additionally, the HC group underwent a thorough examination to ensure that no movement disorder or injury could have significantly affected handwriting.

All participants were asked to complete eight handwriting tasks following a pre-filled template:

1. Drawing an Archimedes spiral;

2. Writing in cursive the letter *l*;
3. The bigram *le*;
4. The trigram *les*;
5. Writing in cursive the word *lektorka* ("female teacher" in Czech);
6. *porovnat* ("to compare");
7. *nepopadnout* ("to not catch");
8. Writing in cursive the sentence *Tramvaj dnes už nepojede* ("The tram won't go today").

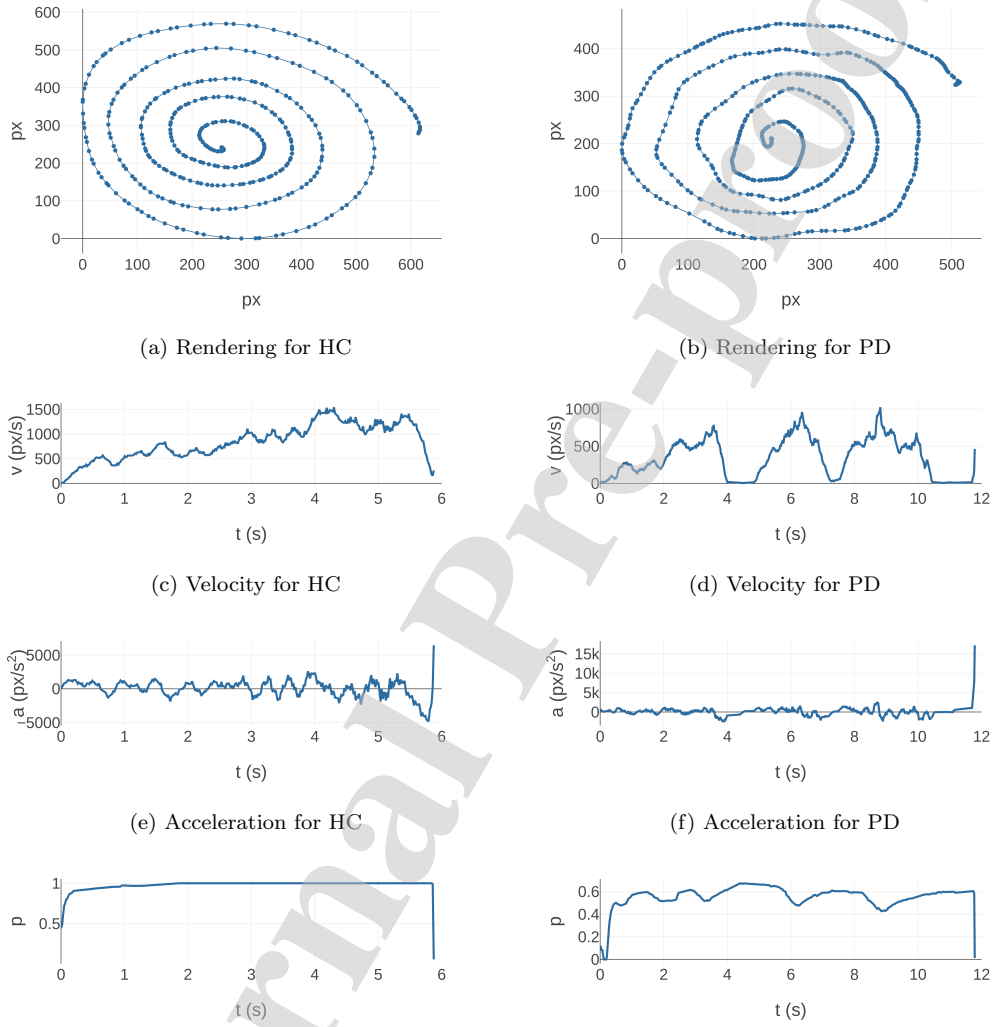
Since not all participants completed each task, we considered only those subjects who completed each of the eight tasks, i.e. 36 PD and 36 HC.

The handwriting signals were recorded using a Wacom Intuos digitizing tablet, overlaid with a blank sheet of paper. Like many other professional tablets, the raw data acquired are the x- and y-coordinates of the pen tip, the corresponding time stamps, measures of pen inclination, i.e. tilt-x and tilt-y, and pen pressure. The button status is also available, which is a binary variable with value 0 for pen-ups ("in-air movement") and 1 for pen-downs ("on-surface movement"). The sampling rate was 200 samples per second. Few sample images of healthy and Parkinsonian writing are depicted in Fig. 1.

3.2. NewHandPD

The NewHandPD database (Pereira, Weber, Hook, Rosa & Papa 2016) is an extension of the previous HandPD corpus (Pereira, Pereira, Silva, Masieiro, Weber, Hook & Papa 2016). The first database consisted of images from two drawing tasks, i.e. the typical spiral cognitive test and a modified spiral ("meander") test performed by healthy individuals and people with Parkinson's disease. However, the new corpus, NewHandPD, contains both offline images and online signals (time-based sequences) of the two groups. The handwriting signals were acquired through a technology other than a tablet, i.e. an electronic smart pen (BiSP).

Specifically, NewHandPD contains images and dynamic data from 31 patients and 35 healthy people. The gender of the participants was fairly balanced (39 males and 29 female), while most of them were right-handed writers (59 of 66 participants). They were asked to complete a handwriting-based test consisting of the following 12 exams:



(a) Rendering for HC (b) Rendering for PD

(c) Velocity for HC (d) Velocity for PD

(e) Acceleration for HC (f) Acceleration for PD

(g) Pressure for HC (h) Pressure for PD

Figure 1: Archimedes spiral task performed by a healthy adult (on the left) and a Parkinsonian patient (on the right). From top to bottom are the rendered task, and the velocity, acceleration and pressure profile (the curves are plotted as a function of advancement during writing). A doctor can typically only see the handwritten pattern left on the paper. Dynamic handwriting features, on the other hand, reveal a lot more information. Parkinsonian handwriting, in fact, shows irregularities and fluctuations that are not shown by the healthy handwriting. (Note that these data do not belong to PaHaW or NewHandPD, but were acquired during the HAND project (Angelillo et al. 2019)).

1. Four tasks related to spirals;
2. Four tasks related to meanders;
3. Two circled movements (one in-air and another on-surface);
4. Two diadochokinesis tests (one with the left hand and the other with the right).

The electronic smart pen recorded the following temporal data in its six channels for each exam:

1. Microphone;
2. Finger grip;
3. Axial pressure of ink refill;
4. Tilt and acceleration in x direction;
5. Tilt and acceleration in y direction;
6. Tilt and acceleration in z direction.

4. Methods

This section introduces our proposed methodology to exploit the potential of sequential information hidden in the time-series handwriting signals for automatic PD identification. Traditionally, medical diagnosis is based on subjective observations from different series of clinical tests. In our study, a computer-aided procedure is proposed to exploit non-visual information for such tests. As an additional element to the medical diagnosis, an objective result is provided, which outperforms the state-of-the-art. Figure 2 depicts a schematic workflow of the proposed method, while details are provided in the following.

4.1. Input Features

It is assumed that raw time-series data sampled from a conventional digitizing tablet are available: pen position, time stamp, pen pressure, pen inclination, and button status. Kinematic and pressure features can be derived from these raw measures. Kinematic features include the tangential, horizontal and vertical displacement, velocity, acceleration, and jerk. Displacement is the straight-line distance between two consecutive sampled points:

$$d_i = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2},$$

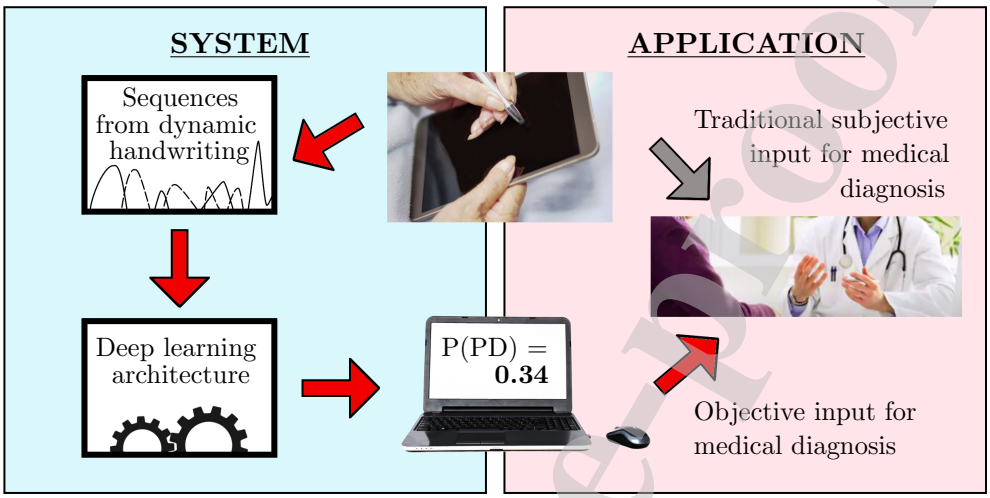


Figure 2: Schematic overview of the proposed system. The data of different handwriting tasks performed by the same subject are acquired through a digitizing tablet. The handwriting specimen of a single task is represented as a multi-dimensional vector of several dynamic features to be provided to a deep learning machine. The system output estimates whether or not the sample belongs to the Parkinsonian class.

where $i = 2, \ldots, Z$ (where Z is the number of sampled points), and $d_1 = 0$. Given the typically high sampling rate of the acquisition device, it generally provides a good approximation of the actual pen trajectory. From this measure, velocity, acceleration, and jerk can be calculated straightforwardly as the first, second, and third derivative of displacement, respectively. This feature set can be enriched by (separately) considering displacement, velocity, acceleration, and jerk along the horizontal and vertical directions. Additionally, to use the pressure data, together with the raw value, we also calculated the first derivative of pressure, which represents the rate of change of pressure over time. An overview of the input features we have considered is provided in Table 1.

These features are suitable for our classification problem, as several studies, e.g. (Broderick et al. 2009, Smits et al. 2014), reported alterations of Parkinsonian handwriting in terms of writing time, writing size, applied pressure, and velocity fluctuations. Note that we do not consider other commonly used spatio-temporal variables, such as stroke size and duration, overall time, etc., as they are expressed as a single-valued feature rather than a time-

Feature	r/d	Description
x	r	x-coordinate of the pen position during handwriting
y	r	y-coordinate of the pen position during handwriting
Pressure	r	Pressure exerted over the writing surface
Tilt-x	r	Angle between the pen and the surface plane
Tilt-y	r	Angle between the pen and the plane vertical to the surface
Button status	r	Boolean variable indicating whether the pen is on-surface or in-air
Displacement	d	Pen trajectory during handwriting
Velocity	d	Rate of change of displacement with respect to time
Acceleration	d	Rate of change of velocity with respect to time
Jerk	d	Rate of change of acceleration with respect to time
Horizontal/vertical displacement	d	Displacement in the horizontal/vertical direction
Horizontal/vertical velocity	d	Velocity in the horizontal/vertical direction
Horizontal/vertical acceleration	d	Acceleration in the horizontal/vertical direction
Horizontal/vertical jerk	d	Jerk in the horizontal/vertical direction
First derivative of pressure	d	Rate of change of pressure with respect to time

Table 1: Dynamic handwriting features. Abbreviations: r = raw feature; d = derived feature.

dependent vector feature (Drotár et al. 2014*a*).

Each handwriting sample S_n can therefore be represented as a multidimensional vector of m dynamic features, where each feature X_i consists of a sequence of T time-steps:

$$S_{n=1}^{N} = \{X_1^n, X_2^n, \dots X_m^n\}$$
$$X_{i=1}^{m} = \{x_i^{t_1}, x_i^{t_2}, \dots x_i^T\},$$

where N is the size of the dataset. The length of the sequential data recorded by the tablet can be arbitrarily long and depends on the time taken by the subject, as well as on the task performed. Since the input sequences can be of varying length for each sample, we fix the time-step length in the pre-processing step. Relatively long time sequences can negatively affect training time, while concise features can lead to underfitting. On the basis of the available data, we first propose to compute the average length of the overall sequences and then to use this length as a cut-off. When the sequences are shorter than the cut-off length, zero-padding is added.

4.2. One-dimensional Convolutions

The time-dependent sequences are fed into one-dimensional (1D) convolutional layers with stride greater than 1. The advantage of employing 1D convolution is two-fold. First, these layers sub-sample the input sequences,

thereby reducing the overall training cost of the RNN model. Second, 1D convolutions can extract local temporal information from the input sequences, thus performing a pre-training step towards learning meaningful temporal dependencies. Combining 1D convolutions and RNNs is beneficial, especially when dealing with very long sequences that would hardly be processed with an RNN. In our case, we can have a few thousands time-steps for a sequence. The effect of the convolutional layers is to turn the long input sequence into much shorter (down-sampled) pieces of higher-level, locally invariant features. The sequence of extracted features then represents the input for the RNN component of the network.

Several filters of varying sizes are used in each layer to extract information across multiple time-scales. In particular, we employed two convolutional layers in cascade. The first applies 8 filters, with a kernel of size 5 and stride 5. The following layer involves a higher number of 16 filters with a reduced kernel of size 3 and stride 3. A commonly used ReLU nonlinearity follows both layers. It is common practice to augment the number of filters in the following layers as the low-level features of the previous one can be combined in several ways to obtain higher-level representations.

4.3. Bidirectional Gated Recurrent Units

Recently, deep learning models such as recurrent neural networks have gained popularity in sequential data analysis (Yu et al. 2019). Unlike a conventional feed-forward neural network, an RNN has a recurrent hidden state h_i^t, whose activation at a given time t depends on the previous state at time $t-1$. This is shown in the following equation:

$$h_i^t = g\left(W.x_i^t + U.h_i^{t-1} + b\right),$$

where W and U are weight matrices, b is the bias term, x_i is the input vector and g the activation function. Despite their effectiveness in modeling sequential data, RNNs are known to suffer from the vanishing gradient problem due to which they may fail to capture long-term dependencies. To address this issue, element-wise non-linearities are typically adopted, which employ two types of recurrent units: the Long-Short Term Memory (LSTM) and the Gated Recurrent Unit (GRU). Although both variants can improve performance, we have chosen to use a GRU-based model as GRUs are less computationally expensive than LSTMs due to the lower number of gates

and therefore fewer parameters to learn.

To further enhance learning, we propose to use Bidirectional GRU (BiGRU) layers. In a BiGRU, two independent GRUs are combined in a bidirectional fashion, with one reading the input sequence in the forward direction. Conversely, the other reads the same sequence in the backward direction. The hidden states from each GRU are then concatenated, as shown in the following:

$$\begin{aligned}
\left(h_i^t\right)_f &= GRU_f\left(x_i^t, h_i^{t-1}\right), \ \ \forall t \in [1, T_i] \\
\left(h_i^t\right)_b &= GRU_b\left(x_i^t, h_i^{t-1}\right), \ \ \forall t \in [T_i, 1] \\
h_i^t &= \left[\left(h_i^t\right)_f; \left(h_i^t\right)_b\right],
\end{aligned}$$

where f and b stand for *forward* and *backward*, respectively. We then process a sequence in both directions to capture patterns that a unidirectional model might overlook.

Two BiGRU layers (32 hidden units each) are stacked on top of the previously mentioned convolutional layers. The two BiGRU layers are interleaved by a conventional dropout and a recurrent dropout, both with a dropout rate of 0.1, to further mitigate overfitting. The output of the last BiGRU is finally sent to an output neuron, with a sigmoid activation attached. Since the problem to be solved can be modeled as a binary classification task (PD/HC), the overall network is required to minimize a classical binary cross-entropy loss function. Training was done using back-propagation with the Adam optimizer and a learning rate of 0.001 on randomly sampled mini-batches of size 16. The overall combined Convolutional-BiGRU model is illustrated in Figure 3.

5. Experiments

In this section, we report the results of a series of experiments aimed at assessing the effectiveness of the proposed method:

- The first experiment evaluated the predictive potential of the proposed system on the PaHaW dataset and ascertained the contribution of the individual subsets of features to the overall classification accuracy;

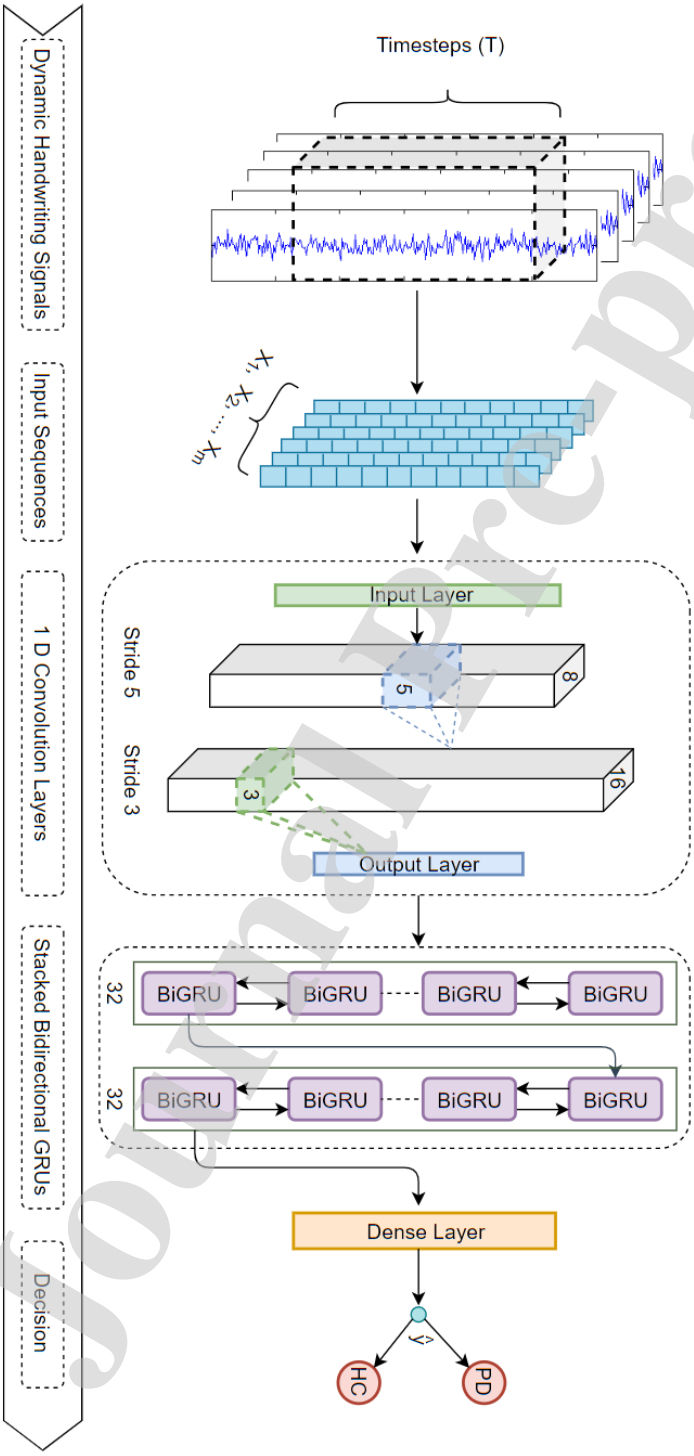


Figure 3: Architectural details of the convolutional-BiGRU model for PD/HC prediction. Raw data from handwriting signals are converted into feature sequences. The input to the model is a sequence of length m and each time-step is a vector of features listed in Table 1. There are two 1D convolutional layers with 8 and 16 filters and strides of 5 and 3, respectively. Two Bi-directional GRU layers, each with 32 units, follow. The output layer has a single neuron with sigmoid activation to predict one of the two classes (PD or HC).

- The second experiment fairly compared, on the same dataset and with the same validation scheme, the proposed method with state-of-the-art approaches to PD detection through dynamic handwriting analysis;
- The third experiment consisted of ablation studies aimed at justifying some architectural choices we made for the construction of the model;
- Finally, the fourth experiment evaluated the model with the best configuration on the NewHandPD dataset to further validate its robustness on data acquired through a slightly different technology.

In the following, the mean accuracy values are reported, averaged over all the iterations of a 10-fold cross-validation scheme. This validation strategy is usually preferred when the size of the data is small. Moreover, for the best model, we also report classification performance in terms of area under the ROC curve (AUC), sensitivity, and specificity, which are commonly used in diagnostic settings.

5.1. Classification Results on PaHaW

Table 2 summarizes the mean accuracy values reported by the proposed model by varying the feature set given as input. Excellent performance is observed in the overall *derived* feature set, including kinematic as well as pressure features, calculated from the raw input acquired by the tablet. The highest predictive potential achieved a mean accuracy of over 90% in almost all cases. In contrast, the overall *raw* feature set reported the lowest classification rates. Not surprisingly, the kinematic features, which contribute most to the aforementioned derived feature set, exhibit the top second accuracy for all tasks among the individual feature groups. These results highlight the effectiveness of these features in capturing the impairments Parkinsonian patients have as they typically do not write with the same constancy as healthy subjects, showing a lower writing speed, with continuous acceleration peaks, e.g. (Kotsavasiloglou et al. 2017, Jerkovic et al. 2019). Significantly lower results, on the other hand, are obtained with pressure features, if considered alone. Patients generally apply less pressure on the writing surface; moreover, the pressure signal assumes erratic values due to muscular difficulties (Rosenblum et al. 2013). However, pressure is generally considered controversial in the literature, especially from the perspective of signature verification (Linden et al. 2018, Diaz et al. 2019*b*), as results differ among studies. Another interesting observation is that pen inclination resulted in

Task	Raw	Inclination	Pressure	Kinematic	Derived
Spiral	70.36%	63.39%	76.25%	85.00%	93.75%
lll	67.50%	87.68%	74.46%	93.75%	96.25%
le le le	71.25%	78.39%	72.68%	92.50%	88.75%
les les les	69.11%	79.11%	65.54%	88.75%	90.00%
lektorka	63.93%	65.54%	61.07%	90.00%	93.75%
porovnat	61.96%	73.21%	68.57%	91.07%	91.25%
nepopadnout	69.11%	78.75%	67.68%	88.57%	92.50%
Sentence	65.89%	80.71%	60.89%	95.00%	92.50%

Table 2: Classification performance of the proposed method. The contribution of individual subsets of features is shown for each task.

relatively better performance, with mean accuracy of more than 80% in two cases. The pen angle information is typically discarded in most related studies. The present results indicate that pen inclination can also be exploited in addition to kinematic and pressure information to further enrich feature representation. It is important to recall that all of these features are first fed to a series of convolutional layers, so the final set of sequences that is provided to the recurrent layers is expected to be a rich and robust representation of the discriminating attributes between PD subjects and healthy controls. These findings also corroborate the hypothesis that sequence learning may be preferred to holistic approaches for PD detection through the dynamics of handwriting.

Further considerations from Table 2 can be drawn by looking at the results obtained task by task. In general, the different feature sets agree that *lll*, *les les les* and the sentence are among the most discriminating tasks. Nonsense words composed of one or more character repetitions have been used frequently for PD assessment, e.g. (Bidet-Ildei et al. 2011, Smits et al. 2014), showing the impairment of Parkinsonian patients in fine motor control during loop-like movements. Indeed, PD patients may produce slower and more irregular movements; moreover, they may write letters in a more segmented fashion, showing micrographia over time when writing. Recently, Senatore & Marcelli (2019) found that Parkinsonian writing during a familiar *l*-shape movement is characterized by a lack of fluency, slowness, and abrupt changes of direction. These difficulties support the hypothesis that the fine-tuning of the motor plan involved is deteriorated due to PD while executing a writing task.

The importance of the sentence task, already observed in (Drotár et al.

2014*b*), is also confirmed in our experiments. In fact, writing a long sentence can require a greater cognitive load, particularly a high degree of simultaneous processing. Therefore, it can increase the effects of the disease on handwriting. The high degree of simultaneous processing is due to several reasons, including the involvement of linguistic skills, attention, and memory. Producing loop-like movements and writing a sentence offers the opportunity to better evaluate the motor plan between one character or word and the next. In fact, a hesitation or pause between two characters or words can highlight the need to re-plan the writing activity. Conversely, fluid writing reveals the presence of early motor planning (Bidet-Ildei et al. 2011, Diaz et al. 2021). In particular, a sentence allows one to capture a large number of in-air movements between components; conversely, a single word could be written without leaving the pen from the writing surface (Drotár et al. 2016).

Another observation concerns the spiral task. As mentioned above, the task was undertaken without any significant impact on classification in previous studies, e.g. (Drotár et al. 2014*b*). Instead, similar to what we previously observed (Diaz et al. 2019*a*, Moetesum et al. 2019), the Archimedes spiral task has achieved high classification accuracy here for almost all feature sets. This reinforces the clinical validity of the task, as clinical experts commonly use it for screening for early signs of PD. One reason may be due to the time it takes to complete the task, as spiral drawing requires continuous on-surface strokes in all directions and, therefore, can better capture changes in the dynamics of handwriting in all directions.

In Table 3, we report the classification performance of the best performing feature set, i.e. derived features, in terms of AUC, sensitivity, and specificity. In general, high values are obtained for all metrics and all tasks, confirming the applicability of the proposed method. There is usually a trade-off between sensitivity and specificity. In the present work, the method appears to be slightly biased in favor of specificity. This suggests that a screening test based on our tool will be better at correctly classifying healthy subjects. To further validate the proposed method, we also illustrate (in Fig. 4) the ROC plots for each of the eight tasks where the highest true positive and lowest false positive rates validate the robust discriminating power of the proposed model.

Task	AUC	Sensitivity	Specificity
Spiral	93.12%	95.00%	92.50%
lll	96.88%	92.50%	100.00%
le le le	92.50%	85.00%	92.50%
les les les	91.88%	92.50%	87.50%
lektorka	91.88%	92.50%	95.00%
porovnat	91.88%	87.50%	95.00%
nepopadnout	96.25%	87.50%	97.50%
Sentence	93.75%	90.00%	95.00%

Table 3: Classification performance of the best performing feature set (derived features) in terms of other well-known metrics.

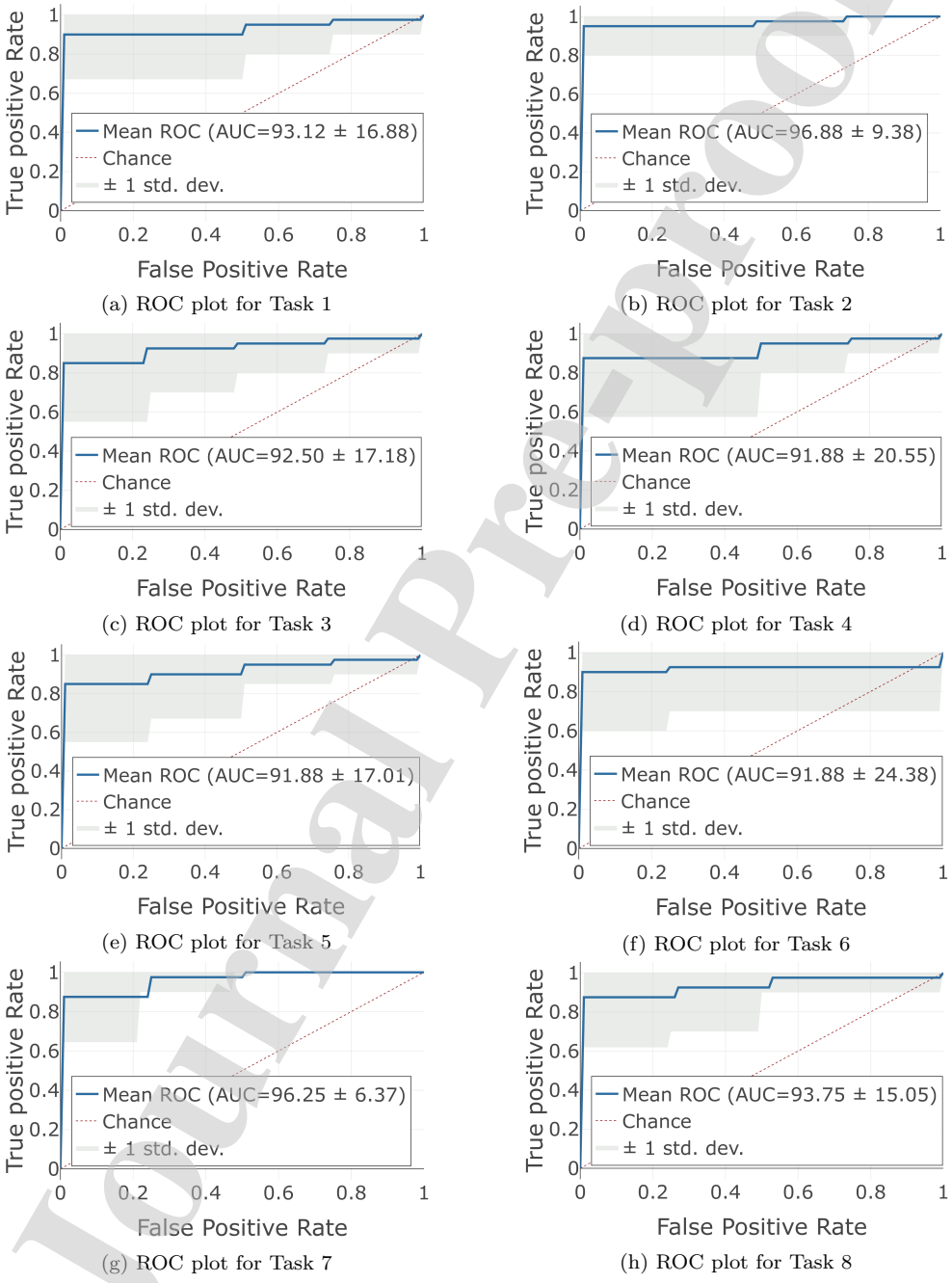
Task	(Drotár et al. 2016)	(Impedovo 2019)	(Angelillo et al. 2019)	(Diaz et al. 2019*a*)	*This work*
Spiral	62.80%	97.33%	53.75%	75.00%	93.75%
lll	72.30%	97.47%	67.08%	64.16%	96.25%
le le le	71.00%	95.12%	72.50%	58.33%	88.75%
les les les	66.40%	93.17%	57.91%	71.67%	90.00%
lektorka	65.20%	96.79%	54.58%	75.41%	93.75%
porovnat	73.30%	95.96%	63.75%	63.75%	91.25%
nepopadnout	67.60%	96.76%	61.67%	70.00%	92.50%
Sentence	76.50%	92.05%	70.42%	67.08%	92.50%

Table 4: Performance comparison with state-of-the-art approaches on PaHaW.

5.2. Comparison with State-of-the-Art on PaHaW

To further establish the effectiveness of the proposed method, a comparative analysis with state-of-the-art approaches aimed at detecting PD through the dynamics of handwriting is presented in Table 4. It should be noted that, during the comparison, it was ensured that the same evaluation metric (mean accuracy) and the same validation scheme (10-fold cross-validation) were used. Indeed, this is the evaluation protocol usually adopted for PaHaW. Furthermore, since different classifiers (support vector machines, random forests, etc.) were used in these studies, reporting different classification accuracies, for a fair comparison, we report here only the best results of these studies for each task.

The first baseline selected for comparison consists of the traditional, handcrafted dynamic features proposed by Drotár et al. (2016). In that work, the horizontal and vertical components of the pen position were segmented into on-surface and in-air strokes as a function of the button status value. Based on this segmentation, kinematic (displacement, velocity, acceleration, and jerk), spatio-temporal (on-surface time) and pressure features were derived. This feature extraction step resulted in either a single-valued feature or a vector feature. For all the resulting vector features, the following basic statistical measures were calculated: mean; median; standard deviation; 1st percentile;



(a) ROC plot for Task 1

(b) ROC plot for Task 2

(c) ROC plot for Task 3

(d) ROC plot for Task 4

(e) ROC plot for Task 5

(f) ROC plot for Task 6

(g) ROC plot for Task 7

(h) ROC plot for Task 8

Figure 4: Task-wise ROC plots for the best feature set.

99th percentile; 99th – 1st percentile. The overall feature vectors were then fed into traditional statistical classifiers, after keeping the most discriminating subset with an *a priori* selection of features made on the overall dataset before cross-validation. It is worth noting that supervised feature selection strategies should be *nested* within the cross-validation iterations, so that the most discriminating features are chosen based only on the training set, while the test set is kept aside. Resorting to an *a priori* selection of features over the entire dataset accidentally introduces a bias in the classification process which may lead to overoptimistic performance. This is well-known in the machine learning community; see, for example, (Hastie et al. 2009).

The second baseline against which we compare the proposed method is the extension of the previously mentioned features proposed by Impedovo (2019). The author has enriched the conventional feature set with additional features obtained by applying the Sigma-Lognormal model, the Maxwell-Boltzmann distribution and the well-known Discrete Fourier Transform to the velocity profile of handwriting. Based on these features, significantly better results were obtained. However, it is worth pointing out that the author used the same *non-nested* validation scheme previously applied by Drotár *et al.* In fact, the author moved from their result as a baseline, then improved it. This means that, while remarkable, the provided contribution still suffers the same feature selection bias as the study of Drotár et al. (2014*b*)

The third baseline consists of the results we obtained by replicating the experiments carried out by Drotár et al. (2014*b*), in which, in order to uncover hidden complexities of handwriting, features based on Shannon and Rényi entropy, signal-to-noise ratio, and empirical mode decomposition (Drotár et al. 2014*b*) were also computed (Angelillo et al. 2019). In the work cited, we used a *nested* feature selection in which the most discriminating features were chosen based only on the training set at each cross-validation iteration. In the work, we showed the detrimental effect on classification accuracy caused by inadvertently introducing the feature selection bias into the machine learning workflow.

Finally, the fourth baseline against which we compared the proposed method consists in the use of features automatically extracted by a 2D Convolutional Neural Network model (Diaz et al. 2019*a*). More specifically, the well-known VGG16 deep network (Simonyan & Zisserman 2014), pre-trained

on ImageNet (Deng et al. 2009), was applied as a feature extractor to multiple "views" of the same static representation of the handwriting. This representation was obtained by embedding dynamic temporal information during the image generation to retain velocity/in-air information. The overall feature vectors were fed into standard statistical classifiers to achieve the final classification, after dynamically retaining features with a *nested* cross-validation scheme.

As can be inferred from Table 4, the sequence learning approach based on 1D convolutions and BiGRUs significantly surpasses, by a considerable margin, the previously proposed procedures based on traditional dynamic features and 2D convolutions on the same dataset. This confirms the effectiveness of the proposed method to serve as a candidate solution for real use in a clinical setting. It is also worth noting that the feature selection bias problem does not apply here. Interestingly, all related works generally agree that the sentence task is among the most discriminating. Conversely, as noted earlier, the spiral task generally reports relatively lower performance when traditional dynamic features are used.

5.3. Ablation Study

We also report the results of ablation studies we carried out to choose the model architecture. These results are summarized in Table 5 and Table 6. It is worth mentioning that these results were obtained by feeding the models with the derived feature set. The first ablation study (Table 5) was conducted to observe the performance of different RNN-based models, which served as a baseline for the outcome of the second ablation study. The features were fed directly into the recurrent units without any convolutional layers. It was observed that the Bidirectional GRU-based model outperformed the other RNN variants on each of the tasks.

The second ablation study (Table 6) mainly concerned the evaluation of the effectiveness of jointly exploiting convolutional and recurrent units. It can be observed that applying convolution on the sequences provided as input, before feeding them into the BiGRU layers, significantly improves classification accuracy. Moreover, it is worth pointing out that, due to the sub-sampling of the time sequences obtained using different stride size (greater than one), the training complexity of all RNN models is also reduced. Overall, these findings further validate our hypothesis that time-based handwriting sequences

Task	BiRNN	BiLSTM	BiGRU
Spiral	84.29%	87.86%	88.57%
lll	81.07%	83.39%	83.57%
le le le	75.71%	75.71%	82.32%
les les les	79.64%	80.00%	84.82%
lektorka	74.11%	75.89%	80.00%
porovnat	77.14%	78.39%	82.32%
nepopadnout	75.54%	82.32%	83.75%
Sentence	83.57%	85.00%	86.25%

Table 5: Comparison between different RNN models without convolution.

Task	BiRNN	BiLSTM	BiGRU
Spiral	88.33%	90.00%	93.75%
lll	91.25%	94.38%	96.25%
le le le	85.00%	88.50%	88.75%
les les les	87.50%	89.67%	90.00%
lektorka	88.75%	92.38%	93.75%
porovnat	87.50%	88.75%	91.25%
nepopadnout	89.40%	91.00%	92.50%
Sentence	90.00%	92.32%	92.50%

Table 6: Comparison between different RNN models with 1D convolution.

contain unique patterns that can be enhanced by convolution and identified by Bidirectional GRUs for PD classification.

5.4. Classification Results on NewHandPD

To validate the generalization capacity of our approach, we ran a series of experiments on NewHandPD. This dataset was not seen during the configuration of our system. Therefore, the best configuration found has been used here.

Specifically, the experiments with NewHandPD have been performed using an experimental protocol similar to that used in (Ribeiro et al. 2019). This included using 65% of the data for training, 10% for validation and 25% for testing. The results obtained have been averaged after repeating the experiment for 20 runs. It is worth noting that we applied the same data preprocessing suggested in (Ribeiro et al. 2019), which consisted of removing outliers by cutting off values below the 5th percentile and above the 90th percentile on each channel. Moreover, a z-score normalization was applied.

The experimental results are provided in Table 7 for each type of exam. Similar to previous results, we reported performance in terms of AUC, sen-

Task	AUC	Sensitivity	Specificity	Accuracy
Spiral	98.25%	90.00%	98.00%	94.44%
Meander	97.75%	90.00%	92.00%	91.11%
$\text{Circle}_{\text{s}}^{\dagger}$	92.25%	85.00%	92.00%	88.89%
$\text{Circle}_{\text{a}}^{\dagger}$	85.91%	85.62%	85.50%	85.56%
$\text{Diadochokinesis}_{\text{R}}$	71.00%	55.00%	78.00%	67.78%
$\text{Diadochokinesis}_{\text{L}}$	73.50%	65.00%	76.00%	71.11%

Circled movements on surface and in the air.

Table 7: Classification performance on NewHandPD following the experimental protocol of (Ribeiro et al. 2019) with our method.

Task	(Pereira et al.) (2016)	(Pereira et al.) (2018)	(Ribeiro et al.) (2019)	*This work*
Spiral	77.53%	78.26%	89.48%	94.44%
Meanders	87.14%	80.75%	92.24%	91.11%
$\text{Circle}_{\text{s}}^{\dagger}$	-	68.04%	-	88.89%
$\text{Circle}_{\text{a}}^{\dagger}$	-	73.41%	-	85.56%
$\text{Diadochokinesis}_{\text{R}}$	-	73.59%	-	67.78%
$\text{Diadochokinesis}_{\text{L}}$	-	76.32%	-	71.11%

Circled movements on surface and in the air.
"-" means results not reported.

Table 8: Performance comparison with state-of-the-art approaches on NewHandPD.

sitivity, specificity and accuracy. Competitive results have been obtained, especially for the spiral-based exam. In addition, the specificity results are consistently greater than sensitivity in all cases. These two effects give consistency to our method as performance with PaWaH also showed similar findings.

Furthermore, we have analyzed the previous literature with this particular database in order to contextualize our results. In (Pereira, Weber, Hook, Rosa & Papa 2016), the authors proposed to model the time-series of the NewHandPD dataset as images. The images were designed using the six available channels as well as the temporal sequences. Then, different CNN architectures were exploited. Specifically, the authors experimented with spirals and meanders. Among the different configurations studied, their best results are shown in Table 8. They correspond to a network pre-trained on ImageNet, accepting 128×128 images and using 75 % of the dataset for training. The authors also reported the results obtained using other classifiers, such as Optimum-Path Forest.

A similar approach was presented in (Pereira et al. 2018). Again, several

CNN architectures were investigated to discriminate between healthy controls and PD patients. In this case, all exams available in NewHandPD were studied using samples from 20 healthy controls and 14 PD patients. The results shown in Table 8 were achieved with 50 % of the specimens for training and using the ImageNet-based network. As a step forward, the authors also presented the results obtained by combining all the exams to achieve a single classification.

For the most relevant comparison, we have selected the work of (Ribeiro et al. 2019), in which stacks of Bidirectional Gated Recurrent Units were employed with an attention layer on top. The authors introduced a bag-of-sampling concept for selecting samples of signal sequences provided in the NewHandPD dataset. The results show that this approach led to better classification outcomes compared to previous studies (Table 8). The experiments were performed using 65% of data for training, 10% for validation and 25% for testing. Instead of using the entire database, the authors used 25 control subjects and 14 patients.

The overall results of our proposed model on the NewHandPD data with a similar experimental protocol show a significant improvement in classification when compared to the results of Pereira, Weber, Hook, Rosa & Papa (2016) and Pereira et al. (2018). When compared with Ribeiro et al. (2019), it is observed that our method has improved results when the spiral is used for classification, while in the case of meanders, our method behaves comparatively with (Ribeiro et al. 2019). It is also noteworthy that our results are computed using all the samples, i.e. 35 healthy and 31 diseased subjects, provided in the NewHandPD database.

Finally, it is worth pointing out that all the results shown correspond to the best configurations studied in each paper for the NewHandPD database. In our work, we have used this database only to demonstrate the generalization capacity of our approach.

6. Conclusion

The growing body of evidence on computerized dynamic handwriting analysis supports the hypothesis that handwriting measures can capture the physical and cognitive characteristics of individuals. In particular, since

handwriting difficulties in Parkinsonian patients have been documented for a long time, such an analysis is promising to help assess Parkinson's disease. The best prospect of this line of research is the integration of new medical tools into current clinical practices to increase the level of diagnostic accuracy. Domain experts can be provided with these easy-to-use, user-friendly tools in their daily practice, without the need for any specific computing expertise. In this sense, a handwriting-based tool represents an attractive choice as it not only provides professionals with a prompt automatic response, but also allows them to store useful metadata related to patient medical records for later use. Of course, handwriting-based decision support tools are not expected to replace standard techniques or humans, but rather provide additional evidence to support their clinical assessment.

In this study, we have proposed a new model based on one-dimensional convolutions and Bidirectional GRUs to identify distinctive patterns in the handwriting sequences of PD patients and controls. Different sets of dynamic features acquired from on-line graphomotor samples of both groups were fed to the model as input. Convolutional layers perform sub-sampling and learn effective feature representations before sending sequences to the Bidirectional GRU part of the network. The results of our experimental study indicate the effectiveness of the proposed technique with respect to the state-of-the-art. The proposed method, in fact, outperformed other "holistic" approaches, thus confirming the effectiveness of the sequence learning paradigm for processing sequential handwriting data. We believe that in addition to the quantitative results, providing a new perspective on the same problem can help clarify some underlying mechanisms still unknown in the future and offer new insights that may be particularly useful for this specific domain. Another observation concerns the exploitation of two datasets whose specimens have been acquired through different technologies. Although previous and recent literature typically used a single dataset for model development and evaluation, the use of a second dataset helped us confirm the robustness of the proposed method.

A significant limitation of the present study is the small size of the datasets we employed, which can somewhat influence the generalizability of the results obtained. Unfortunately, developing a large benchmark dataset is still one of the major open issues in the pattern recognition community working in this field (Vessio 2019, Faundez-Zanuy et al. 2020). This applies not

only to PD, but also to other neurodegenerative disorders (Impedovo et al. 2019). Nevertheless, despite these constraints, the reported performance values are indeed very promising and the results of this study are expected to make way for a working system in the clinical settings.

Authorship Contribution Statement

Moises Diaz: Conceptualization, Validation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Momina Moetesum:** Methodology, Software, Investigation, Resources, Writing - Review & Editing. **Imran Siddiqi:** Conceptualization, Investigation, Writing - Review & Editing, Supervision. **Gennaro Vessio:** Conceptualization, Validation, Writing - Original Draft, Writing - Review & Editing.

Declaration of 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.

Acknowledgment

Part of this research was funded by the Higher Education Commission (HEC), Pakistan, under grant number 8910/Federal/NRPU/R&D/HEC/2017. Additional funding was received from the Spanish government's MIMECO TEC2016-77791-C4-1-R and PID2019-109099RB-C41 research projects and the European Union FEDER program/funds.

References

Afonso, L. C., Rosa, G. H., Pereira, C. R., Weber, S. A., Hook, C., Albuquerque, V. H. C. & Papa, J. P. (2019), 'A recurrence plot-based approach for Parkinson's disease identification', *Future Generation Computer Systems* **94**, 282–292. https://doi.org/10.1016/j.future.2018.11.054.

Ali, L., Zhu, C., Zhou, M. & Liu, Y. (2019), 'Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection', *Expert Systems with Applications* **137**, 22–28. https://doi.org/10.1016/j.eswa.2019.06.052.

Ammour, A., Aouraghe, I., Khaissidi, G., Mrabti, M., Aboulem, G. & Belahsen, F. (2020), 'A new semi-supervised approach for characterizing the arabic on-line handwriting of Parkinson's disease patients', *Computer Methods and Programs in Biomedicine* **183**, 104979. https://doi.org/10.1016/j.cmpb.2019.07.007.

Angelillo, M. T., Impedovo, D., Pirlo, G. & Vessio, G. (2019), Performance-driven handwriting task selection for Parkinson's disease classification, *in* 'International Conference of the Italian Association for Artificial Intelligence', Springer, pp. 281–293. https://doi.org/10.1007/978-3-030-35166-3_20.

Ascherio, A. & Schwarzschild, M. A. (2016), 'The epidemiology of Parkinson's disease: risk factors and prevention', *The Lancet Neurology* **15**(12), 1257–1272. https://doi.org/10.1016/S1474-4422(16)30230-7.

Bhat, S., Acharya, U. R., Hagiwara, Y., Dadmehr, N. & Adeli, H. (2018), 'Parkinson's disease: Cause factors, measurable indicators, and early diagnosis', *Computers in Biology and Medicine* **102**, 234–241. https://doi.org/10.1016/j.compbiomed.2018.09.008.

Bidet-Ildei, C., Pollak, P., Kandel, S., Fraix, V. & Orliaguet, J.-P. (2011), 'Handwriting in patients with Parkinson disease: effect of L-dopa and stimulation of the sub-thalamic nucleus on motor anticipation', *Human Movement Science* **30**(4), 783–791. https://doi.org/10.1016/j.humov.2010.08.008.

Broderick, M. P., Van Gemmert, A. W., Shill, H. A. & Stelmach, G. E. (2009), 'Hypometria and bradykinesia during drawing movements in individuals with Parkinson's disease', *Experimental Brain Research* **197**(3), 223–233. https://doi.org/10.1007/s00221-009-1925-z.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. & Bengio, Y. (2014), Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv:1406.1078.

Danna, J., Velay, J.-L., Eusebio, A., Véron-Delor, L., Witjas, T., Azulay, J.-P. & Pinto, S. (2019), 'Digitalized spiral drawing in Parkinson's disease:

A tool for evaluating beyond the written trace', *Human Movement Science* **65**, 80–88. https://doi.org/10.1016/j.humov.2018.08.003.

De Stefano, C., Fontanella, F., Impedovo, D., Pirlo, G. & di Freca, A. S. (2019), 'Handwriting analysis to support neurodegenerative diseases diagnosis: A review', *Pattern Recognition Letters* **121**, 37–45. https://doi.org/10.1016/j.patrec.2018.05.013.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. & Fei-Fei, L. (2009), Imagenet: A large-scale hierarchical image database, *in* '2009 IEEE Conference on Computer Vision and Pattern Recognition', IEEE, pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848.

Diaz, M., Ferrer, M. A., Carmona, C. & Plamondon, R. (2021), Improving handwritten signatures fluency via the lognormality principles, *in* R. Plamondon, A. Marcelli & M. A. Ferrer, eds, 'The Lognormality Principle and its Applications', World Scientific, pp. 41–63. https://doi.org/10.1142/97898112268300002.

Diaz, M. et al. (2019*a*), 'Dynamically enhanced static handwriting representation for Parkinson's disease detection', *Pattern Recognition Letters* **128**, 204–210. https://doi.org/10.1016/j.patrec.2019.08.018.

Diaz, M. et al. (2019*b*), 'A perspective analysis of handwritten signature technology', *ACM Computing Surveys (CSUR)* **51**(6), 1–39. https://doi.org/10.1145/3274658.

Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z. & Faundez-Zanuy, M. (2014*a*), 'Analysis of in-air movement in handwriting: A novel marker for Parkinson's disease', *Computer Methods and Programs in Biomedicine* **117**(3), 405–411. https://doi.org/10.1016/j.cmpb.2014.08.007.

Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z. & Faundez-Zanuy, M. (2014*b*), 'Decision support framework for Parkinson's disease based on novel handwriting markers', *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **23**(3), 508–516. https://doi.org/10.1109/TNSRE.2014.2359997.

Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z. & Faundez-Zanuy, M. (2016), 'Evaluation of handwriting kinematics and pressure

for differential diagnosis of Parkinson's disease', *Artificial Intelligence in Medicine* **67**, 39–46. https://doi.org/10.1016/j.artmed.2016.01.004.

Eichhorn, T., Gasser, T., Mai, N., Marquardt, C., Arnold, G., Schwarz, J. & Oertel, W. (1996), 'Computational analysis of open loop handwriting movements in Parkinson's disease: a rapid method to detect dopamimetic effects', *Movement Disorders: Official Journal of the Movement Disorder Society* **11**(3), 289–297. https://doi.org/10.1002/mds.870110313.

Faundez-Zanuy, M., Fierrez, J., Ferrer, M. A., Diaz, M., Tolosana, R. & Plamondon, R. (2020), 'Handwriting biometrics: Applications and future trends in e-security and e-health', *Cognitive Computation* **12**(5), 940–953. https://doi.org/10.1007/s12559-020-09755-z.

Ferrer, M. A., Diaz, M., Carmona-Duarte, C. & Plamondon, R. (2020), 'iDeLog: iterative dual spatial and kinematic extraction of sigma-lognormal parameters', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **42**(1), 114–125. https://doi.org/10.1109/TPAMI.2018.2879312.

Hastie, T., Tibshirani, R. & Friedman, J. (2009), *The elements of statistical learning: data mining, inference, and prediction*, Springer Science & Business Media. https://doi.org/10.1007/978-0-387-84858-7.

Hochreiter, S. & Schmidhuber, J. (1997), 'Long short-term memory', *Neural Computation* **9**(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Impedovo, D. (2019), 'Velocity-based signal features for the assessment of Parkinsonian handwriting', *IEEE Signal Processing Letters* **26**(4), 632–636. https://doi.org/10.1109/LSP.2019.2902936.

Impedovo, D., Pirlo, G., Vessio, G. & Angelillo, M. T. (2019), 'A handwriting-based protocol for assessing neurodegenerative dementia', *Cognitive Computation* **11**(4), 576–586. https://doi.org/10.1007/s12559-019-09642-2.

Jerkovic, V. M., Kojic, V., Miskovic, N. D., Djukic, T., Kostic, V. S. & Popovic, M. B. (2019), 'Analysis of on-surface and in-air movement in

handwriting of subjects with Parkinson's disease and atypical parkinsonism', *Biomedical Engineering/Biomedizinische Technik* **64**(2), 187–194. https://doi.org/10.1515/bmt-2017-0148.

Kotsavasiloglou, C., Kostikis, N., Hristu-Varsakelis, D. & Arnaoutoglou, M. (2017), 'Machine learning-based classification of simple drawing movements in Parkinson's disease', *Biomedical Signal Processing and Control* **31**, 174–180. https://doi.org/10.1016/j.bspc.2016.08.003.

Linden, J., Marquis, R., Bozza, S. & Taroni, F. (2018), 'Dynamic signatures: A review of dynamic feature variation and forensic methodology', *Forensic science international* **291**, 216–229. https://doi.org/10.1016/j.forsciint.2018.08.021.

Lipton, Z. C., Berkowitz, J. & Elkan, C. (2015), A critical review of recurrent neural networks for sequence learning. arXiv:1506.00019.

Moetesum, M., Siddiqi, I., Javed, F. & Masroor, U. (2020), Dynamic handwriting analysis for Parkinson's disease identification using C-BiGRU model, *in* '17th International Conference on Frontiers in Handwriting Recognition (ICFHR)', IEEE. (To Appear).

Moetesum, M., Siddiqi, I., Vincent, N. & Cloppet, F. (2019), 'Assessing visual attributes of handwriting for prediction of neurological disorders–A case study on Parkinson's disease', *Pattern Recognition Letters* **121**, 19–27. https://doi.org/10.1016/j.patrec.2018.04.008.

Parisi, L., RaviChandran, N. & Manaog, M. L. (2018), 'Feature-driven machine learning to improve early diagnosis of Parkinson's disease', *Expert Systems with Applications* **110**, 182–190. https://doi.org/10.1016/j.eswa.2018.06.003.

Pereira, C. R., Pereira, D. R., Rosa, G. H., Albuquerque, V. H., Weber, S. A., Hook, C. & Papa, J. P. (2018), 'Handwritten dynamics assessment through convolutional neural networks: An application to Parkinson's disease identification', *Artificial Intelligence in Medicine* **87**, 67–77. https://doi.org/10.1016/j.artmed.2018.04.001.

Pereira, C. R., Pereira, D. R., Silva, F. A., Masieiro, J. P., Weber, S. A., Hook, C. & Papa, J. P. (2016), 'A new computer vision-based approach to

aid the diagnosis of parkinson's disease', *Computer Methods and Programs in Biomedicine* **136**, 79–88.

Pereira, C., Weber, S., Hook, C., Rosa, G. & Papa, J. (2016), Deep learning-aided parkinson's disease diagnosis from handwritten dynamics, *in* '2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)', Ieee, pp. 340–346.

Randhawa, B. K., Farley, B. G. & Boyd, L. A. (2013), 'Repetitive transcranial magnetic stimulation improves handwriting in Parkinson's disease', *Parkinson's Disease* **2013**. https://doi.org/10.1155/2013/751925.

Ribeiro, L. C., Afonso, L. C. & Papa, J. P. (2019), 'Bag of samplings for computer-assisted Parkinson's disease diagnosis based on recurrent neural networks', *Computers in Biology and Medicine* **115**, 103477. https://doi.org/10.1016/j.compbiomed.2019.103477.

Rios-Urrego, C. D., Vásquez-Correa, J. C., Vargas-Bonilla, J. F., Nöth, E., Lopera, F. & Orozco-Arroyave, J. R. (2019), 'Analysis and evaluation of handwriting in patients with Parkinson's disease using kinematic, geometrical, and non-linear features', *Computer Methods and Programs in Biomedicine* **173**, 43–52. https://doi.org/10.1016/j.cmpb.2019.03.005.

Rosenblum, S., Samuel, M., Zlotnik, S., Erikh, I. & Schlesinger, I. (2013), 'Handwriting as an objective tool for Parkinson's disease diagnosis', *Journal of Neurology* **260**(9), 2357–2361. https://doi.org/10.1007/s00415-013-6996-x.

Senatore, R. & Marcelli, A. (2019), 'A paradigm for emulating the early learning stage of handwriting: Performance comparison between healthy controls and Parkinson's disease patients in drawing loop shapes', *Human Movement Science* **65**, 89–101. https://doi.org/10.1016/j.humov.2018.04.007.

Simonyan, K. & Zisserman, A. (2014), Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556.

Smits, E. J., Tolonen, A. J., Cluitmans, L., Van Gils, M., Conway, B. A., Zietsma, R. C., Leenders, K. L. & Maurits, N. M. (2014), 'Standardized handwriting to assess bradykinesia, micrographia and tremor in

Parkinson's disease', *PloS one* **9**(5), e97614. https://doi.org/10.1371/journal.pone.0097614.

Teulings, H.-L. & Stelmach, G. E. (1991), 'Control of stroke size, peak acceleration, and stroke duration in Parkinsonian handwriting', *Human Movement Science* **10**(2-3), 315–334. https://doi.org/10.1016/0167-9457(91)90010-U.

Vessio, G. (2019), 'Dynamic handwriting analysis for neurodegenerative disease assessment: A literary review', *Applied Sciences* **9**(21), 4666. https://doi.org/10.3390/app9214666.

You, Q., Jin, H., Wang, Z., Fang, C. & Luo, J. (2016), Image captioning with semantic attention, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition', pp. 4651–4659. https://doi.org/10.1109/CVPR.2016.503.

Yu, Y., Si, X., Hu, C. & Zhang, J. (2019), 'A review of recurrent neural networks: LSTM cells and network architectures', *Neural Computation* **31**(7), 1235–1270. https://doi.org/10.1162/neco_a_01199.

Zham, P., Kumar, D. K., Dabnichki, P., Poosapadi Arjunan, S. & Raghav, S. (2017), 'Distinguishing different stages of Parkinson's disease using composite index of speed and pen-pressure of sketching a spiral', *Frontiers in Neurology* **8**, 435. https://doi.org/10.3389/fneur.2017.00435.

Zhang, X.-Y., Yin, F., Zhang, Y.-M., Liu, C.-L. & Bengio, Y. (2017), 'Drawing and recognizing chinese characters with recurrent neural network', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **40**(4), 849–862. https://doi.org/10.1109/TPAMI.2017.2695539.

Research Highlights:

- Sequence-based dynamic handwriting analysis for Parkinson's Disease identification
- Combination of one-dimensional convolutions and Bi-GRU layers for classification
- Robust feature learning through convolutions on features derived from handwriting
- Experiments on the PaHaW and NewHandPD datasets achieve state-of-the-art results

Credit Author Statement

Authorship Contribution Statement

Moises Diaz: Conceptualization, Validation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Momina Moetesum:** Methodology, Software, Investigation, Resources, Writing - Review & Editing. **Imran Siddiqi:** Conceptualization, Investigation, Writing - Review & Editing, Supervision. **Gennaro Vessio:** Conceptualization, Validation, Writing - Original Draft, Writing - Review & Editing.

***Declaration of Interest Statement**

Declaration of interests

☒ 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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: