



On the use of histograms of oriented gradients for tremor detection from sinusoidal and spiral handwritten drawings of people with Parkinson's disease

João Paulo Folador¹ · Maria Cecilia Souza Santos¹ · Luiza Maire David Luiz¹ · Luciane Aparecida Pascucci Sande de Souza² · Marcus Fraga Vieira³ · Adriano Alves Pereira¹ · Adriano de Oliveira Andrade¹

Received: 1 May 2020 / Accepted: 22 December 2020 / Published online: 7 January 2021

© International Federation for Medical and Biological Engineering 2021

Abstract

Parkinson's disease (PD), whose cardinal signs are tremor, rigidity, bradykinesia, and postural instability, gradually reduces the quality of life of the patient, making early diagnosis and follow-up of the disorder essential. This study aims to contribute to the objective evaluation of tremor in PD by introducing and assessing histograms of oriented gradients (HOG) to the analysis of handwriting sinusoidal and spiral patterns. These patterns were digitized and collected from handwritten drawings of people with PD ($n = 20$) and control healthy individuals ($n = 20$). The HOG descriptor was employed to represent relevant information from the data classified by three distinct machine-learning methods (random forest, k-nearest neighbor, support vector machine) and a deep learning method (convolutional neural network) to identify tremor in participants with PD automatically. The HOG descriptor allowed for the highest discriminating rates (accuracy 83.1%, sensitivity 85.4%, specificity 80.8%, area under the curve 91%) on the test set of sinusoidal patterns by using the one-dimensional convolutional neural network. In addition, ANOVA and Tukey analysis showed that the sinusoidal drawing is more appropriate than the spiral pattern, which is the most common drawing used for tremor detection. This research introduces a novel and alternative way of quantifying and evaluating tremor by means of handwritten drawings.

Keywords Histograms of oriented gradients · Archimedean spiral · Sinusoidal drawing · Tremor · Parkinson's disease

1 Introduction

Parkinson's disease (PD) is a neurodegenerative disorder affecting an area called substantia nigra in the basal ganglia. PD produces a progressive loss of dopaminergic neurons that causes various motor symptoms such as tremor, rigidity, bradykinesia, postural instability, shuffling gait, and non-motor

symptoms such as depression, sleep problems, loss of cognitive function, nerve pain, and intestinal constipation [1].

PD affects about 1% of the over 60-year-old world population. Moreover, according to the World Health Organization (WHO), by 2050, almost two billion people worldwide are expected to be over 60 years old. Thus, 20 million people may be suffering from PD in the future [2].

PD, unfortunately, remains cureless and its diagnosis is not simple. The patient must be assessed by means of standardized clinical exams, diagnostic imaging exams, and the dopaminergic therapy response. However, the cardinal signs of the disease, such as tremors, rigidity, bradykinesia, and postural instability, are features that can mark the presence of the disorder [1, 3, 4].

An adequate understanding of these cardinal signs can help to diagnose the condition. Thus, the evaluation of the hand tremor is a crucial stage of the clinical assessment of the patient with PD, and it can be done by analyzing drawings that can capture tremors. This evaluation may be conducted by a

✉ João Paulo Folador
jpfolador@gmail.com

¹ Centre for Innovation and Technology Assessment in Health, Postgraduate Program in Electrical and Biomedical Engineering, Faculty of Electrical Engineering, Federal University of Uberlândia, Uberlândia, Brazil

² Applied Physiotherapy Department, Federal University of Triângulo Mineiro, Uberaba, Brazil

³ Bioengineering and Biomechanics Laboratory, Federal University of Goiás, Goiânia, Brazil

specialist or computer-based method [1]. Computational methods have the main advantage that they rely on techniques that guarantee consistency and reproducibility.

Currently, several studies demonstrate the use of techniques to extract information from handwritten drawings in real-time. Smits et al. [5] used a tablet and digitizer pen to draw shapes such as circles, stars, spirals, and characters (e.g., “elel”). The authors quantify the tremor, bradykinesia, and micrographia using the writing dynamics (i.e., speed, time, writing size, and frequency) to detect people with PD from a control group (CG).

Features extracted from tablet-based spiral drawings were correlated with part III of the Unified Parkinson Disease Rating Scale (UPDRS) score [6]. The study introduced by Almeida et al. [7] used a digitizing tablet to collect spiral drawings to analyze the correlation between physiological tremor and aging. Westin et al. [8] proposed a new web software for viewing and comparing spiral drawing assessments. Several research groups [9–11] have used Archimedes' spiral drawings to assess hand tremor in PD patients via tablet-based data collection.

Recently, deep learning methods, which have a higher computational cost, have been used to analyze spiral-handwritten drawings. Khatamino et al. [12] reported a convolutional neural network (CNN) with the highest accuracy of 88%. Similarly, Moetesum et al. [13] reported an accuracy of 83% while discriminating PD from a CG group.

Pereira et al. [14] used a smartpen with several attached sensors to obtain data from spiral and meander drawings. The smartpen signals were transformed into pictures, and the images were inputted in CNN to build knowledge. CNN reached an accuracy of 83.7% in the classification of PD and CG.

Tolonen et al. [15] used a tablet and a pen with an attached gyroscope to collect data from drawings such as spirals, circles, and zigzag triangles. They classified PD from other movement disorders and achieved the following accuracy: 82% (essential tremor), 69.8% (functional tremor), and 72.2% (physiological tremor). Similarly, the work presented by Matsumoto et al. [16] used a tablet and a digital pen with a 3-axis accelerometer to evaluate spiral drawings of patients with PD and essential tremor (ET).

Although many studies have been employing inertial sensors and digitizing tablets, there is a lack of research that analyzes digitized drawings made with paper and pen. This type of analysis is relevant because of the simplicity of data collection and its broad availability in the clinical environment.

Kraus and Hoffmann [17] used paper and pencil to collect spiral drawings and then used these images to analyze tremor amplitude. The regression analysis revealed a significant association (88.9%) between the Bain et al. [18] rating scale and tremor amplitude.

Bajaj et al. [19] collected spiral drawings using paper and pen to distinguish cases of tremulous PD from those clinical-

SWEEDs (parkinsonian phenotype with normal presynaptic dopaminergic imaging). The authors assessed the clinical tremor severity (TS), spiral diameter 3-turns (3TD), and spiral density (SD). The sensitivity and specificity in predicting the correct classification were, respectively, 62.5% and 65.0% for TS, 75% and 56.7% for 3TD, and 30.4% and 82.5% for SD.

Pereira et al. [20] used paper and pen to collect spiral drawings from CG and patients with PD. They reached $78.9 \pm 3.5\%$ using the Naïve Bayes classifier. The work presented by Pereira et al. [21] used CNN to extract features from spiral and meanders of handwritten to identify subjects with PD and showed average overall accuracy for meanders of 79.62% and spirals of 89.55%. Passos et al. [22] reached 96% of accuracy to identify people with PD using a complex structure of a deep neural network (DNN) (ResNet-50) to learn the patterns and extract features from the image spiral drawings, and then fed the Optimum-Path Forest (OPT).

In addition, Gupta and Chanda [23] proposed a Fourier transform-based distance, specific to spiral drawings, to extract features from offline images. The authors discriminated patients with PD from a healthy group with an accuracy of 81.66% by using support vector machine.

Archimedes' spiral drawing has been widely used to assess tremor. However, the characteristics and methods used in the analysis of Archimedes' spiral cannot be applied to other types of drawings. Besides, there is a lack of studies comparing the results from Archimedes' spiral with other drawings for tremor quantification. As presented by Daroff et al. [24], handwriting drawings, e.g., sinusoidal patterns, should be employed for the evaluation of tremor. In contrast to a spiral drawing, which captures the hand movement in the same position, the sinusoidal drawing requires the person to slide the hand from one point to another. This difference in the motion pattern may bring essential characteristics to be analyzed.

Folador et al. [25] introduced the analysis of sinusoidal handwritten drawings for tremor evaluation. The authors showed that it was possible to discriminate drawings made from people with PD from a CG by employing computer-vision techniques to extract features from images. A random forest classifier was applied to these characteristics, and overall mean accuracy of 70% was reached.

The present research extends the investigation of Folador et al. [25] by collecting and analyzing a data set with 960 handwritten drawings of spiral and sinusoidal patterns. Twenty people with PD and 20 healthy subjects participated in the experimental trials. The presented results consider the clinical evaluation of tremor by three distinct examiners and the comparative performance of four classifiers in distinguishing the tremor symptom in people with PD from the control group.

The histograms of oriented gradient (HOG) descriptor used in computer vision has been used primarily in human body detection [26] but extended to tsunami victim detection [27],

texture classification [28], traffic sign detection [29], and mammographic image classification [30]. In this research, HOG was introduced for tremor detection in people with PD from the spiral and sinusoidal handwritten drawings. The HOG features were fed into three classical machine-learning techniques, which have been employed in several related studies, i.e., random forest [31–35], k-nearest-neighbor [33, 36, 37], and support vector machine [20, 23, 27, 33–35, 38]. Additionally, the HOG features were fed in a deep learning classifier (convolutional neural network) [12–14, 21].

The main purpose of the proposed method was to employ solely paper and pencil to quantify tremor through spiral and sinusoidal handwriting images by HOG feature extraction. Practical advantages of this method are the following: (i) it can be widely employed as it depends only on the availability of paper and pencil; (ii) it does not need an experimental environment with supervision of drawing parameters such as speed, time, and pressure; (iii) it can be extended to the analysis of distinct types of drawing patterns as it does not depend on specific metrics of a particular kind of drawing pattern.

2 Materials and methods

This transversal research was approved by the Research Ethics Committee of the Federal University of Uberlândia (CAAE 07075413.6.0000.5152). The participants were informed about the data collection procedures and signed a consent form before data collection.

2.1 Participants

Data were collected from people with PD (PwPD) and healthy individuals (i.e., control group—CG). MDS-UPDRS (Unified Parkinson's Disease Rating Scale modified by the Movement Disorder Society in 2008) [39] was applied to PwPD. The inclusion criteria in the control group were the absence of neurological disorder and any physical impairment that might prevent the individual from executing the experimental tasks.

The research included 20 PwPD, as shown in Table 1. They were treated with antiparkinsonian drugs. The experiment occurred while the participants were in the period “ON” of the medication to get a more stable handwritten drawing.

The control group consisted of 20 matched individuals of the same age and sex-related to PwPD, as indicated in Table 1. The statistical equivalence of the groups was verified by the *t* test in which the null hypothesis (H_0) is that the means of PwPD and CG groups are equal, and if it is rejected (*p* value < 0.05), the means are different. The *t* test is commonly used in data that follow a normal distribution. Thus, the normality was confirmed by Shapiro–Wilk test (H_0 : the sample has a normal distribution; if *p* value < 0.05, the null hypothesis is rejected) and the inspection on the quartile-quartile plot (QQ-Plot) [40].

2.2 Procedure for data collection

During the experiment, the participants were asked to use the dominant hand to draw a spiral and a sinusoidal pattern. All the participants were blind evaluated by three physiotherapists with experience in Parkinson's disease. Besides guaranteeing the absence of problems or severe comorbidities that could affect the dominant hand, they also applied the MDS-UPDRS part III (specific questions 3.4, 3.5, 3.6, 3.15, 3.16, and 3.17) to evaluate the hand movement, in particular the presence and severity of hand tremor. These specific items of the MDS-UPDRS were applied to the right and left hand.

The participants were asked to draw a counter-clockwise pattern on a printed drawing before doing the experimental task, as shown in Fig. 1a and b. The purpose of this step was to provide prior knowledge of the experimental task to the participant. Following this, they were asked to freely draw six patterns of each drawing on an A4 white sheet of paper using a 6B black pencil from Faber Castel.

Data collection was performed in two different experimental sessions with an interval of 1 week between them, accounting for the effect of data variability in the analysis and the increase of data samples from a single participant. In total, each participant drew 12 patterns in an experimental session, being six spirals and six sinusoidal. There was no clinical complication with the participants in this period and no change in medication.

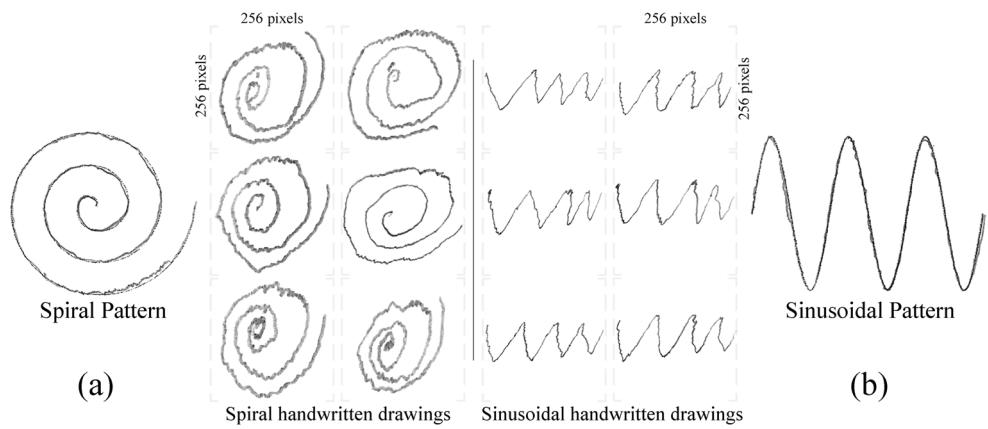
Therefore, a database of 960 images was created with 480 spirals and 480 sinusoidal for both sessions: 240 spirals and 240 sinusoidal images for each experimental session, being half of them collected from PwPD and the other half from

Table 1 Characterization of the research groups

Group	Number of participants	Age in years (mean \pm SD)	Sex (M/F)	Shapiro-Wilk		<i>t</i> test		
				W	<i>p</i> value	<i>t</i> value	degree of freedom	<i>p</i> value
PwPD	20	67.65 \pm 9.10	17/3	0.9517	0.3938	0.2166	37.7580	0.8297
CG	20	67.05 \pm 8.40	17/3	0.9512	0.3853			

PwPD people with Parkinson's disease, CG control group, SD standard deviation, M/F male/female

Fig. 1 **a** The spiral and **b** the sinusoidal patterns drawn by the participants. The pattern drawings were counterbalanced while the other images were produced freely



CG. Hence, a balanced dataset was set up with the same number of samples per class [41].

The drawings were digitized and preprocessed so that features could be estimated. In the last stage, the extracted features were classified using distinct machine-learning methods to discriminate people with tremor from those without visible tremulous activity. Figure 2 illustrates the main steps of the study.

2.3 Computational and data processing environment

The pencil drawings were digitized in 300 dots per inch (dpi) of resolution using a scanner (HP Deskjet 3516 multifunctional). Since all the drawings of each participant were in a single sheet of paper, it was necessary to select individual drawings of each pattern after digitization, using an image manipulation program (GIMP). Each drawing was rescaled to a standard size of 256 by 256 pixels (wide and height) in 96 dpi.

Data processing and all other experiments were carried out in a machine with Intel Core i7 2.40 GHz, with 8 GB DDR3 RAM, 256 SSD of hard disk, and a 2 GB NVIDIA GeForce GT 650 video card on Microsoft Windows 7 Pro 64 bits.

Python programming language 3.6.5 was used with Tensorflow 2.1 (the core of an open-source library for Machine Learning) and Keras 2.3.1 (a deep learning framework), and the

Scientific Python Development Environment (Spyder 3.3.2) was used for coding. R Studio 1.1.456 was used for statistical analyses as well as for graphical visualization.

2.4 Clinical analysis of tremor

In this study, three examiners applied the MDS-UPDRS to PwPD. The scores of the MDS-UPDRS part III (questions 3.4, 3.5, 3.6, 3.15, 3.16, and 3.17) given by each examiner were summed for each participant. The scores of the left and right hands were computed. The score range was between 0 and 48 points.

The analysis of the agreement between examiners was carried out by applying Kendall's coefficient to compare the agreement among all examiners. Kendall's coefficient is a non-parametric statistic and can be used to measure the agreement among several evaluators assessing a given set of n subjects. Its value ranges from 0 (no agreement) to 1 (complete agreement) [42].

2.5 Feature extraction

The histograms of oriented gradients (HOG) proposed and detailed in [26, 28, 43] is a method based on evaluating the

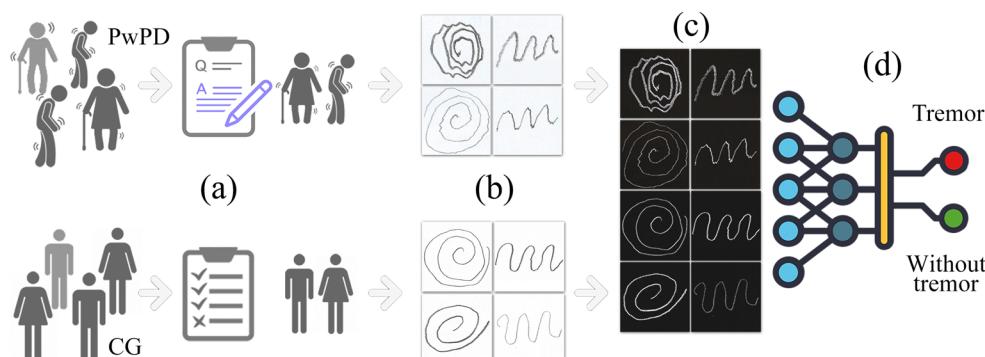


Fig. 2 Flowchart depicting the main stages of the study. **a** Recruitment and clinical assessment of healthy volunteers and people with Parkinson's disease. **b** The handwritten drawings were collected, digitized, and

preprocessed. **c** Features of the digitized images were extracted. **d** The set of features was classified with the aim of discriminating people with tremor from those without visible tremulous activity

normalized local histograms in a dense grid that uses gradient magnitude and angle information for object detection. The distribution of local intensity gradients can represent the appearance and shape of the region analyzed.

Figure 3 illustrates the main steps involved in the estimate of the HOG. In this study, a window of the digitized drawing was divided into small regions named cells, e.g., the blue square in Fig. 3b, of 16 (width) by 16 (height) pixels, and a cell was represented by a vector of one dimension of histogram of gradients. In this study, the histogram of a cell has 9 orientation bins (Fig. 3b), as suggested in [26, 43]. To the normalization process, the block of cells (the green grid in Fig. 3b) was set to 2 by 2 cells.

For estimating the HOG descriptor, the gradients g_x and g_y must be calculated as in Eq. 1, where x and y are the pixel positions in the image and f is the pixel intensity. Therefore, the horizontal target pixel (g_x) intensity is obtained by the difference between the right and left pixel values from it. In the vertical direction, g_y is calculated by the difference between the top and bottom values of the neighbor pixels; an example of the gradients obtained is in Fig. 3c [26, 28].

$$\nabla f(x, y) = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} f(x+1, y) - f(x-1, y) \\ f(x, y+1) - f(x, y-1) \end{bmatrix} \quad (1)$$

The HOG is obtained as the combination of the local histograms of g_x and g_y . The parameters that define a local histogram are the following: the magnitude g in Eq. 2 and the edge orientation θ in Eq. 3 [27, 28].

$$g = \sqrt{g_x^2 + g_y^2} \quad (2)$$

$$\theta = \arctan(g_y / g_x) \quad (3)$$

The number of bins in local histograms was configured to 9, and as the value of θ ranged from 0 to 180° , the first bin ranged from 0 to 20° , the second from 20 to 40° , and so on. After that, the voting process [26] selects the bin based on the value of θ and then adds the pixel magnitude to the bin (Fig. 3d).

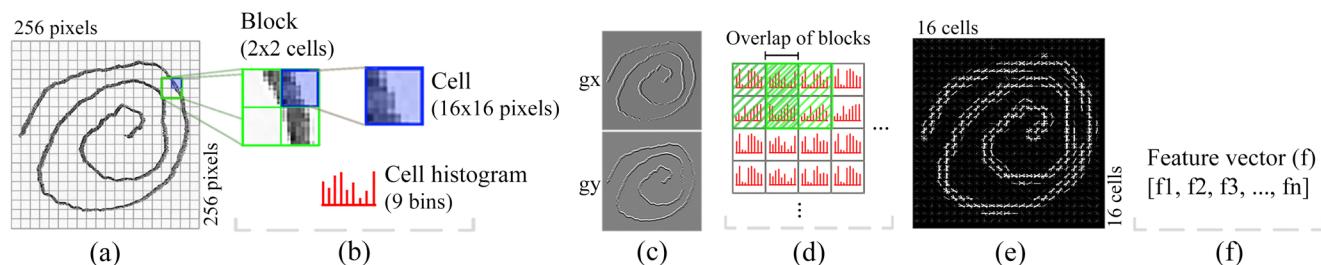


Fig. 3 Basic steps involved in the HOG descriptor estimation: **a** input image, **b** HOG parameters setup such as the image division in cells and blocks, **c** the gradients g_x and g_y are computed, **d** the histograms of each

cell are estimated, and the normalization is processed by block, **e** the HOG descriptor, and **f** a vector of the normalized cell histograms from all blocks are produced.

For improving the invariance of shadows and illumination into the process, the measure of the local histogram is calculated within larger regions named blocks, set as 2 by 2 cells. The block scans the dense grid of the cell histograms produced from left to right and from top to bottom, and the block overlap (stride) is fixed at half of the block size (Fig. 3d). Each block can be normalized using, for instance, L1-norm ($f = v / (\|v\|_1 + c)$) where v is the non-normalized descriptor vector of a block, $\|v\|_1$ is the 1-norm, and c is a constant value ($c = 1$) that prevents division by zero [26]. Each value of the cell histogram is divided by the block normalization value. When the block overlaps, the normalization process repeats. Each of the cells is represented in the final feature vector several times but normalized by different blocks. However, these redundancies increase the performance of the descriptor [26, 28].

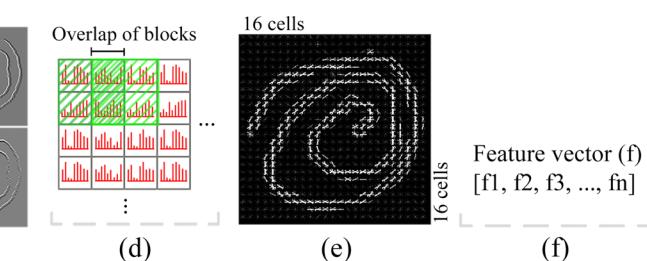
The normalized block descriptors are referred to as histogram of oriented gradient (HOG) descriptors (Fig. 3e). The size of the final feature vector can be calculated multiplying the number of bins (9), the number of cells per block (4), the amount of horizontal overlapping (15), and the amount of vertical overlap (15). In this study, the size of the HOG feature vector, as represented in Fig. 3f, is 8100 (a dimensionality reduction of 87.6% when compared to the original image, which is 65,536). The features used in the test set of each classifier can be visualized in the “Results” section.

2.6 Data classification analysis

The set of HOG feature vectors was classified so that it was possible to discriminate drawings with tremor against those without tremulous activity. Though all PwPD had tremor in this study, it may happen that the individual did not present the symptom during the experimental trials. Likewise, some healthy people could present tremor because of anxiety or other stressful factors during the experiment.

Four supervised classifiers were employed: random forest classifier (RFC), k-nearest-neighbor (KNN), support vector machine (SVM), and convolutional neural network (CNN).

RFC is a type of supervised machine learning algorithm based on ensemble learning. This characteristic allows us to combine different algorithms or the same algorithm to create a



cell are estimated, and the normalization is processed by block, **e** the HOG descriptor, and **f** a vector of the normalized cell histograms from all blocks are produced.

more efficient prediction model. The combination of multiple decision-tree algorithms was used [31]. In general, an RFC takes N objects from the database and builds a decision tree with this data, and every tree predicts the category of the items belonging to it. Finally, the new object is assigned to the class that wins the majority vote [31–33]. In this work, 500 trees were used, such as in [34].

K-nearest-neighbor (KNN) is a type of data classification algorithm that attempts to classify which category the data point is in by looking at the data points around it. It is a non-parametric method generally used for classification and regression [33]. In this work, k was set to 7 neighbors ($k = 7$) as in [36].

SVM is a type of supervised machine learning classification algorithm. The algorithm aims to find a boundary that divides the data in such a way that the misclassification error can be minimized. The nearest points from the decision boundary that maximize the distance between the decision boundary and the points are called support vectors. The decision boundary in support vector machines is called the maximum margin classifier or the maximum margin hyperplane [38]. In this work, a linear kernel was used, as in [26].

A common CNN (two-dimensional or 2D CNN) is a type of deep learning method typically used in the classification of images. Unlike conventional multilayer perceptron architectures, CNN performs the so-called convolution and pooling layers, attempting to reduce the image to its basic features for understanding and classifying it [12, 14, 44].

In this proposal, the HOG features are used to feed a one-dimensional CNN (1D CNN), which is commonly used for sequence processing. In this context, the convolutional layer uses a kernel to extract local patches (subsequences) from the original sequence of features and feeds a fully connected layer to compute the classification [44]. In this study, 1D CNN was configured with two convolutional layers with kernel length of 5, two pooling layers, and with 3 fully connected layers (the first 2 with 16 units and the output layer with 1 unit) to classify PwPD or CG. The configuration was based on [44] and empirically improved.

A 2D CNN was used to extract the features and classify the original images of the database (without HOG) to compare the results. The 2D CNN arrangement was set up with 5 convolutional layers with kernel length of 3 by 3, 5 pooling layers, and with 2 fully connected layers (the first layer with 512 units and the output layer with 1 unit) to classify PwPD and CG. The configuration was based on [44] and empirically improved.

Cross-validation stratified k-fold was used for the evaluation of all the classifiers. This method splits randomly the data into k equally sized groups or folds preserving the same percentage of samples in each class, and then, one group is used to test and the others to train the classifier [45]. In this work,

the 5-fold was employed [34, 46, 47], which means the data are trained/tested 5 times in each experiment.

The training/test was performed separately for session 1 and session 2 for each type of drawing (spirals and sinusoidal waveforms) on a balanced dataset with the same number of samples in each class [41].

The following metrics were employed to assess the quality of the classification results:

- Accuracy (ACC) is the proportion of correct prediction of a given condition as defined in Eq. 4 [30, 46, 48].

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (4)$$

Thus, TP is the number of true positives, TN the number of true negatives, FP the number of false positives, and FN the number of false negatives.

- Sensitivity (SEN) is the measure of the true positive rate, as defined in Eq. 5 [46, 48].

$$\text{SEN} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

- Specificity (SPE) is related to the true negative rate, as defined in Eq. 6 [46, 48].

$$\text{SPE} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (6)$$

- Receiver operating characteristic (ROC) curve presents on the y-axis the values of sensitivity and in the x-axis the complement of specificity (1—specificity) [30, 48].
- The area under the curve (AUC) is given by Eq. 7. The higher the AUC is in a range from 0 to 1, the better the model is at distinguishing between individuals with and without tremor symptoms [48].

$$\text{AUC} = \int_0^1 \text{ROC}(t) \, dt \quad (7)$$

The motor fluctuation was assessed by considering data collection in two different experimental sessions. Thus, the

following models were created for each classifier considering the distinct types of experimental conditions:

- M_1 : a model of spiral drawings of the experimental session 1;
- M_2 : a model of spiral drawings of the experimental session 2;
- M_3 : a model of spiral drawings of the experimental sessions 1 and 2;
- M_4 : a model of sinusoidal drawings of the experimental session 1;
- M_5 : a model of sinusoidal drawings of the experimental session 2;
- M_6 : a model of sinusoidal drawings of the experimental sessions 1 and 2;

The accuracy of the models (M_1 – M_6) was tested for data normality verification. The Shapiro-Wilk test was used to test the null hypothesis that a sample came from a normally distributed population. If p value < 0.05, the null hypothesis (H_0) was rejected. Bartlett's test was applied to verify that the samples had equal variances (H_0). If p < 0.05, the test presents no equality of variances. Finally, the Bonferroni outlier test was used to verify if there was the presence of outliers; on the null hypothesis, the outliers do not differ from the rest of the observations. If p value < 0.05, there is the presence of outliers [49].

The accuracy of the models was assessed by analysis of variance (ANOVA) and Tukey-Kramer to understand the differences of each model. ANOVA is a statistical method widely used to explain variations between two or more group means. The null hypothesis describes no differences between the group means ($H_0: \mu_1 = \mu_2 = \dots = \mu_m$). Suppose ANOVA results in significant differences (H_0 is rejected, p value < 0.05), Tukey is applied for performing multiple pairwise comparison between the means of groups, and the means that are significantly different from each other are highlighted [49, 50].

The analysis of the accuracy of distinct models helps to understand and determine whether there is motor fluctuation [51] between different data collection sessions and also to identify the most suitable type of drawing (i.e., spiral or sinusoidal pattern) for tremor evaluation.

3 Results

3.1 Statistics for control and PD groups

The QQ-plot of the age distribution for each group (PD and CG) was inspected to confirm the normality of the variables [49]. Furthermore, the Shapiro-Wilk normality test [49] confirmed that the distribution of the variable age of both groups was normal ($W = 0.9517$ and p value = 0.3938 for PD; and $W = 0.9512$ and p value = 0.3853 for the CG).

The t test was applied to verify possible differences between the age of the groups [40, 49]. The estimated t test statistic was 0.2166, with 37.758 degrees of freedom and a p value of 0.8297, which is larger than 0.05, meaning that the null hypothesis that there are no significant differences between the ages of PwPD and CG should not be rejected.

People with PD were evaluated by three experienced evaluators in a blind procedure, and they had an average sum of the score of 15.57 to the first session and 16.42 to the second session, with a range of 0–48, indicating that PwPD are at the presence of a slight to mild level of tremor.

Furthermore, the estimated Kendall's coefficient was 0.6610 and the p value 0.0065, which means the coefficient values are significantly different from zero. Therefore, the null hypothesis that the evaluators may disagree with was rejected.

3.2 HOG feature visualization

The vector of HOG features is calculated by multiplying the number of bins, the number of cells per block, and the amount of horizontal and vertical overlapping. In this research, the feature vector had 8100 elements. Figure 4 illustrates these HOG features in a 3D plot. The data are from the test set of PD and CG groups. The visual inspection of the features in Fig. 4 allowed for identifying distinct feature magnitude and variability for the spiral and sinusoidal images. For this reason, three equally sized regions delimited by these features were defined as region 1 (feature 0 to 2700), region 2 (feature 2700 to 5400), and region 3 (feature 5400 to 8100).

The mean and coefficient of variation, together with their respective 95% confidence interval, were estimated for each region, group (PD and CG), and proposed model (M_1 – M_6). The statistics and their confidence interval were estimated by Bootstrap, which is a statistical method for estimating the sampling distribution of a statistic (e.g., mean, coefficient of variation) by sampling with replacement from the original sample [49]. In this research, random sampling was executed 1000 times, as suggested in [49].

Figure 5 shows the statistics mean and coefficient of variation, together with their 95% confidence interval, estimated through Bootstrap. The statistics were computed for HOG features of the spiral and sinusoidal images. In the graphs, the x-axis labels represent the name of the model concatenated with the group and the region delimited by HOG features. For example, the label $M2PD3$ represents model 2 (M_2) of the PD group and region 3 (features from 5400 to 8100). Figure 5a and c show statistics for the features estimated from spiral images, while b and d show statistic values for the sinusoidal images.

In Fig. 5, the gray area highlights the comparison between PD and CG groups of the same model and region. It is possible to note that, in general, there is no overlap between confidence intervals of distinct groups, and when overlaps occur, they are small; for instance, in Fig. 5a, the upper limit of the confidence

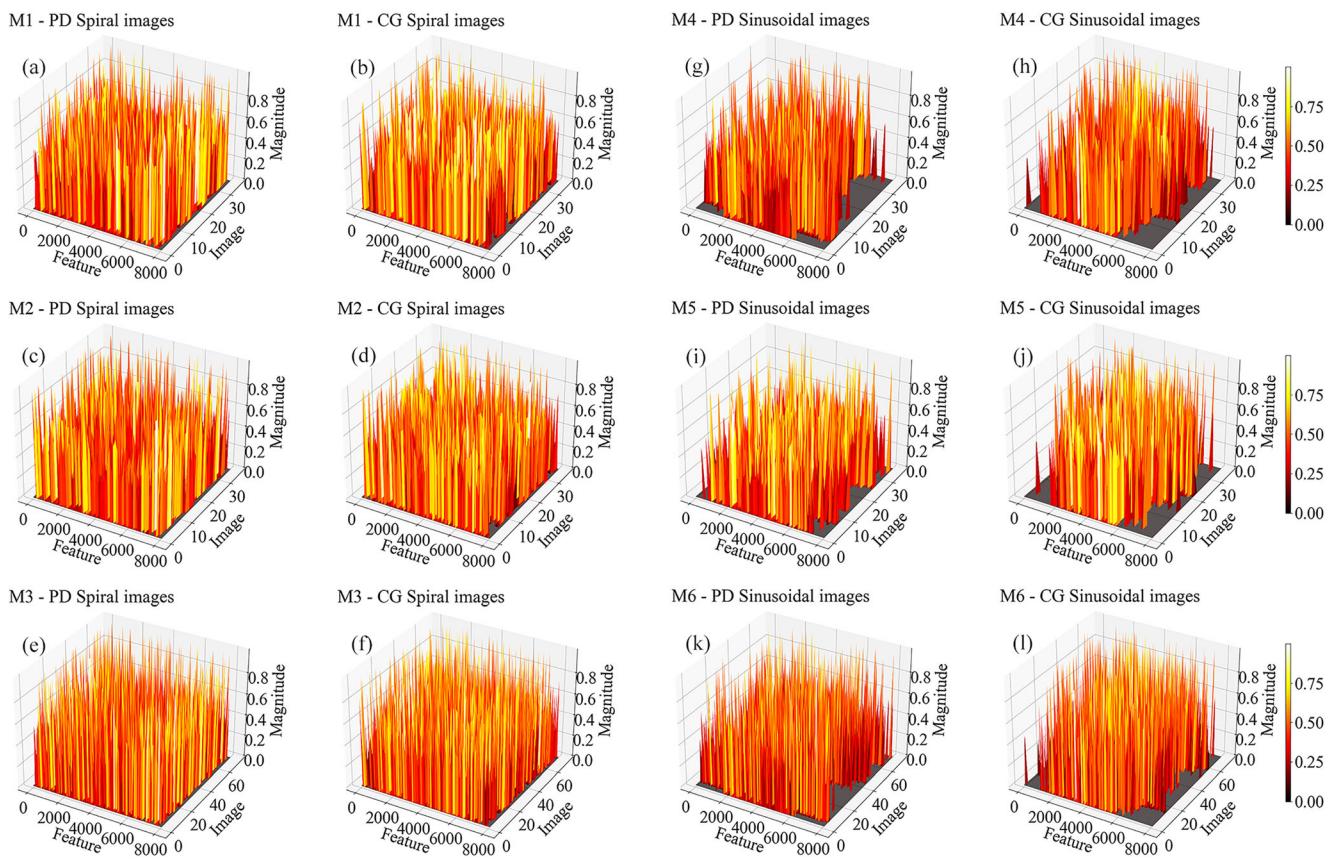


Fig. 4 Visualization of features for PD and CG groups for the test sets used to evaluate each model. Data (36 images per group) from experimental session 1 (**a**, **b**, **g**, and **h**) and experimental session 2 (**c**, **d**, **i**, and **j**) are presented. In addition, data (72 images per group) of sessions 1 and 2 (**e**, **f**, **k**, and **l**) are shown

interval for the mean of M3CG3 (0.02001) is equal to the lower limit of M3PD3. The results shown in Fig. 5 show that the confidence intervals for the statistics are narrow, suggesting an accurate estimate, for example, in Fig. 5d, the estimates for M6CG2 (coefficient of variation 3.20853, CI 3.20778, 3.20928) do not overlap with those of M6PD2 (coefficient of variation 3.18846, CI 3.1877, 3.18921). The results shown in Fig. 5 highlight differences between feature values in distinct regions (from 1 to 3), which can also be observed in Fig. 4 for each group and type of image.

3.3 Classification results

3.3.1 Random forest classifier

Table 2 shows the results of each session and type of drawing. The italicized results are the highest mean accuracy, sensitivity, and specificity, suggesting that data from both experimental sessions and the sinusoidal pattern are more relevant for the objective evaluation of hand tremor.

Figure 6 shows the ROC curve and the AUC value calculated for each subset. The results for RFC-M₆ confirm the most accurate RFC model shown in Table 2.

3.3.2 K-nearest-neighbor

Table 3 shows the KNN results of each session and type of drawing. The italicized outcomes in the table are the highest mean accuracy, sensitivity, and specificity, implying that data from both experimental sessions and the sinusoidal drawing are more relevant for the objective evaluation of hand tremor.

Figure 7 shows the ROC curve and the AUC value calculated for each subset. The results for KNN-M₆ confirm the identification of the most accurate KNN model shown in Table 3.

3.3.3 Support vector machine

Table 4 shows the results of each session and the type of drawing from the SVM classifier. The italicized results are

Fig. 5 **a**, **b** Mean HOG and its 95% confidence interval for spiral and sinusoidal images, respectively. **c**, **d** Coefficient of variation of HOG and its 95% confidence interval for the spiral and sinusoidal images, respectively. The statistics are presented for distinct models (M1 to M6), groups (PD and CG), and regions delimited by features (1, 2, and 3). For instance, M1CG1 is the statistic for model 1, group CG, and region 1. The 95% confidence interval was relatively narrow and hence difficult to see in the figure.

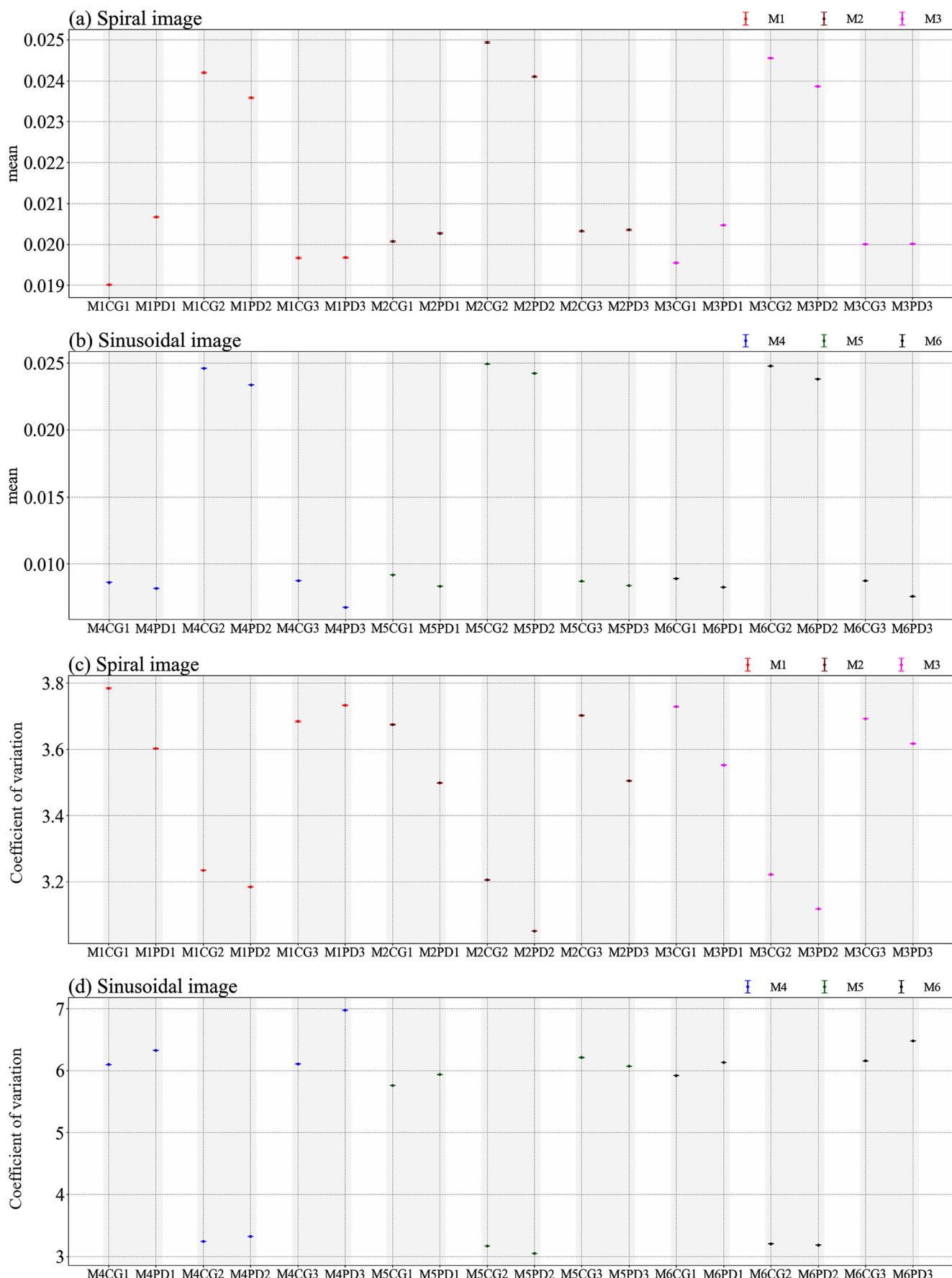


Table 2 Random forest classifier (RFC)

Model	Session	Type of drawing	Accuracy (ACC)			Sensitivity (SEN)			Specificity (SPE)		
			Min	Max	Mean ± SD	Min	Max	Mean ± SD	Min	Max	Mean ± SD
M ₁	1	Spiral	0.625	0.771	0.700 ± 0.057	0.542	0.833	0.717 ± 0.103	0.625	0.792	0.683 ± 0.062
M ₂	2	Spiral	0.708	0.854	0.754 ± 0.016	0.667	0.833	0.758 ± 0.055	0.667	0.833	0.750 ± 0.059
M ₃	1, 2	Spiral	0.708	0.760	0.735 ± 0.022	0.583	0.875	0.767 ± 0.101	0.646	0.833	0.704 ± 0.068
M ₄	1	Sinusoidal	0.750	0.854	0.800 ± 0.036	0.708	0.917	0.792 ± 0.074	0.708	0.875	0.808 ± 0.062
M ₅	2	Sinusoidal	0.708	0.875	0.783 ± 0.061	0.667	0.833	0.783 ± 0.067	0.583	0.917	0.783 ± 0.113
M ₆	1, 2	Sinusoidal	0.730	0.875	0.817 ± 0.047	0.708	0.896	0.808 ± 0.061	0.688	0.875	0.825 ± 0.070

These results are from a test set

SD standard deviation

the highest mean accuracy, sensitivity, and specificity, suggesting that data from both experimental sessions and the sinusoidal pattern are more relevant for the objective evaluation of hand tremor.

Finally, Fig. 8 illustrates the ROC curve and the AUC value calculated for each subset to the SVM classifier. The results for SVM-M₆ confirm the identification of the most accurate SVM model demonstrated in Table 4.

3.3.4 One-dimensional convolutional neural network

Table 5 shows the results of each session and the type of drawing from the 1D CNN classifier. The italicized results are the highest mean accuracy, sensitivity, and specificity, suggesting that data from both experimental sessions and the sinusoidal pattern are more relevant for the objective evaluation of hand tremor.

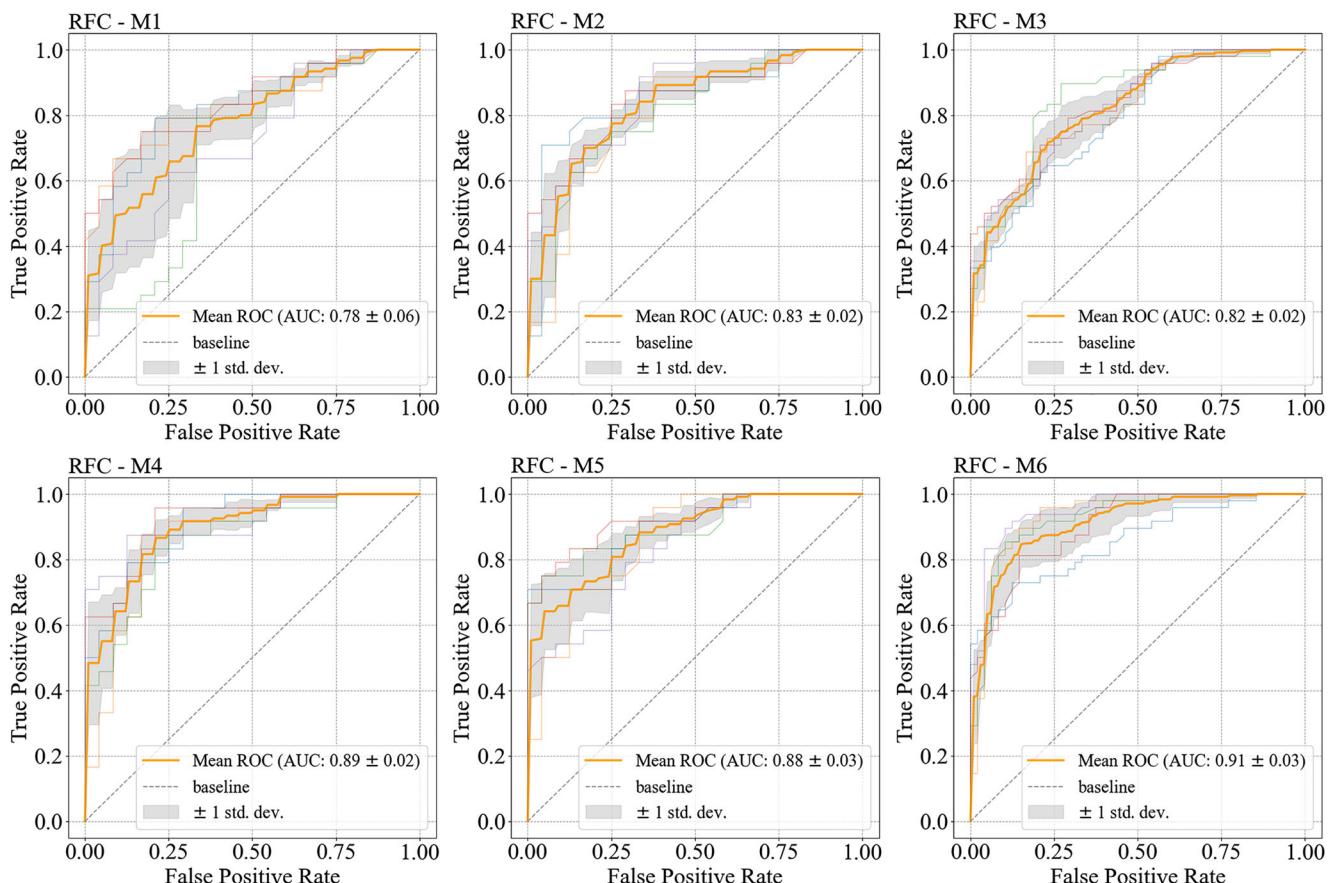


Fig. 6 ROC curve and AUC values of the RFC model. The M₁, M₂, and M₃ graphs represent the results of the spiral drawing test set of the data collected from session 1, session 2, and all data together, respectively.

Similarly, M₄, M₅, and M₆ show the results of the sinusoidal drawings on the test set of the data collected from session 1, session 2, and all data together

Table 3 k-Nearest Neighbor (KNN)

Model	Session	Type of drawing	Accuracy (ACC)			Sensitivity (SEN)			Specificity (SPE)		
			Min	Max	Mean ± SD	Min	Max	Mean ± SD	Min	Max	Mean ± SD
M ₁	1	Spiral	0.667	0.729	0.700 ± 0.028	0.625	0.750	0.675 ± 0.041	0.583	0.833	0.725 ± 0.089
M ₂	2	Spiral	0.646	0.750	0.696 ± 0.036	0.667	0.792	0.750 ± 0.053	0.542	0.750	0.642 ± 0.077
M ₃	1, 2	Spiral	0.698	0.802	0.742 ± 0.035	0.646	0.813	0.738 ± 0.061	0.708	0.792	0.746 ± 0.028
M ₄	1	Sinusoidal	0.750	0.854	0.788 ± 0.036	0.792	0.958	0.842 ± 0.061	0.667	0.750	0.733 ± 0.033
M ₅	2	Sinusoidal	0.688	0.854	0.750 ± 0.063	0.750	0.917	0.842 ± 0.061	0.500	0.792	0.658 ± 0.096
M ₆	1, 2	Sinusoidal	0.750	0.865	0.804 ± 0.041	0.792	0.896	0.854 ± 0.037	0.625	0.833	0.754 ± 0.077

These results are from a test set

SD standard deviation

Figure 9 illustrates the ROC curve and the AUC value calculated for each subset to the 1D CNN classifier. The results for 1D CNN-M₆ confirm the identification of the most accurate model shown in Table 5.

3.3.5 Two-dimensional convolutional neural network

Table 6 shows the results of each session and the type of drawing from the 2D CNN classifier. The italicized results

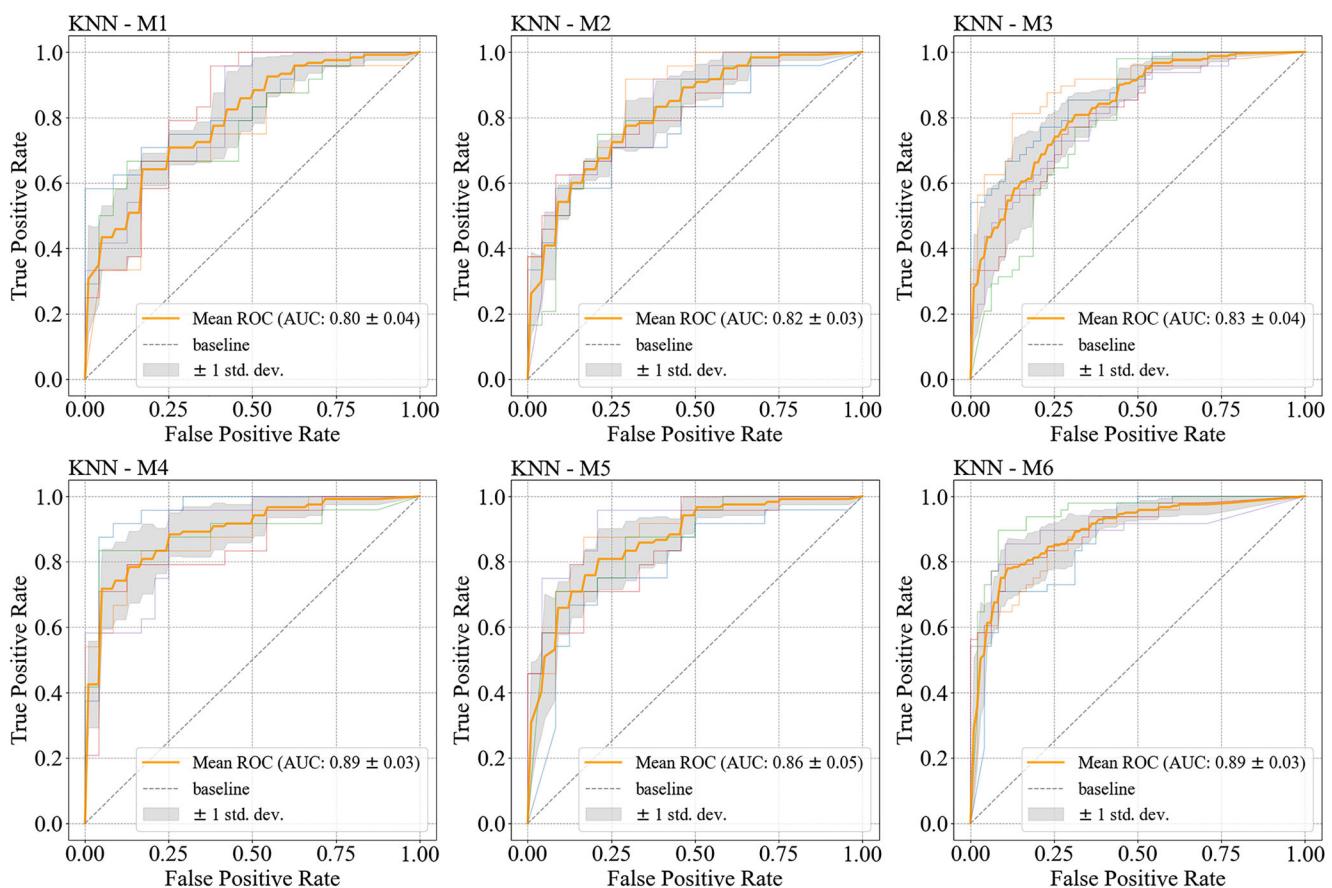


Fig. 7 ROC curve and AUC values of the KNN model. The M₁, M₂, and M₃ graphs represent the results of spiral drawing test set of the data collected from session 1, session 2, and all data together, respectively.

Similarly, M₄, M₅, and M₆ show the results of the sinusoidal drawings on the test set of the data collected from session 1, session 2, and all data together

Table 4 Support vector machine (SVM)

Model	Session	Type of drawing	Accuracy (ACC)			Sensitivity (SEN)			Specificity (SPE)		
			Min	Max	Mean ± SD	Min	Max	Mean ± SD	Min	Max	Mean ± SD
M ₁	1	Spiral	0.667	0.750	0.696 ± 0.028	0.583	0.750	0.667 ± 0.070	0.625	0.792	0.725 ± 0.082
M ₂	2	Spiral	0.563	0.854	0.696 ± 0.104	0.500	0.875	0.717 ± 0.127	0.500	0.833	0.675 ± 0.113
M ₃	1, 2	Spiral	0.667	0.760	0.696 ± 0.033	0.542	0.833	0.704 ± 0.100	0.563	0.813	0.687 ± 0.080
M ₄	1	Sinusoidal	0.729	0.833	0.783 ± 0.045	0.708	0.833	0.783 ± 0.049	0.708	0.875	0.783 ± 0.055
M ₅	2	Sinusoidal	0.688	0.875	0.783 ± 0.061	0.667	0.875	0.800 ± 0.072	0.708	0.875	0.767 ± 0.062
M ₆	1, 2	Sinusoidal	0.781	0.875	0.823 ± 0.032	0.708	0.958	0.817 ± 0.082	0.792	0.875	0.829 ± 0.031

These results are from a test set

SD standard deviation

are the highest mean accuracy, sensitivity, and specificity, suggesting that data from both experimental sessions and the sinusoidal pattern are more relevant for the objective evaluation of hand tremor.

Finally, Fig. 10 illustrates the ROC curve and the AUC value calculated for each subset to the 2D CNN classifier. The results for 2D CNN-M₆ confirm the identification of the most accurate model shown in Table 6.

3.3.6 Evaluation of models

Table 7 displays the data normality (Shapiro-Wilk test), variance (Bartlett's test), the presence of outliers (Bonferroni Outlier test), and the analysis of variance (ANOVA) that were employed to compare the accuracy of each model (M₁–M₆) for each classifier (RFC, KNN, SVM, 1D CNN, and 2D CNN). Thus, the appropriate ANOVA fitting shall be fulfilled

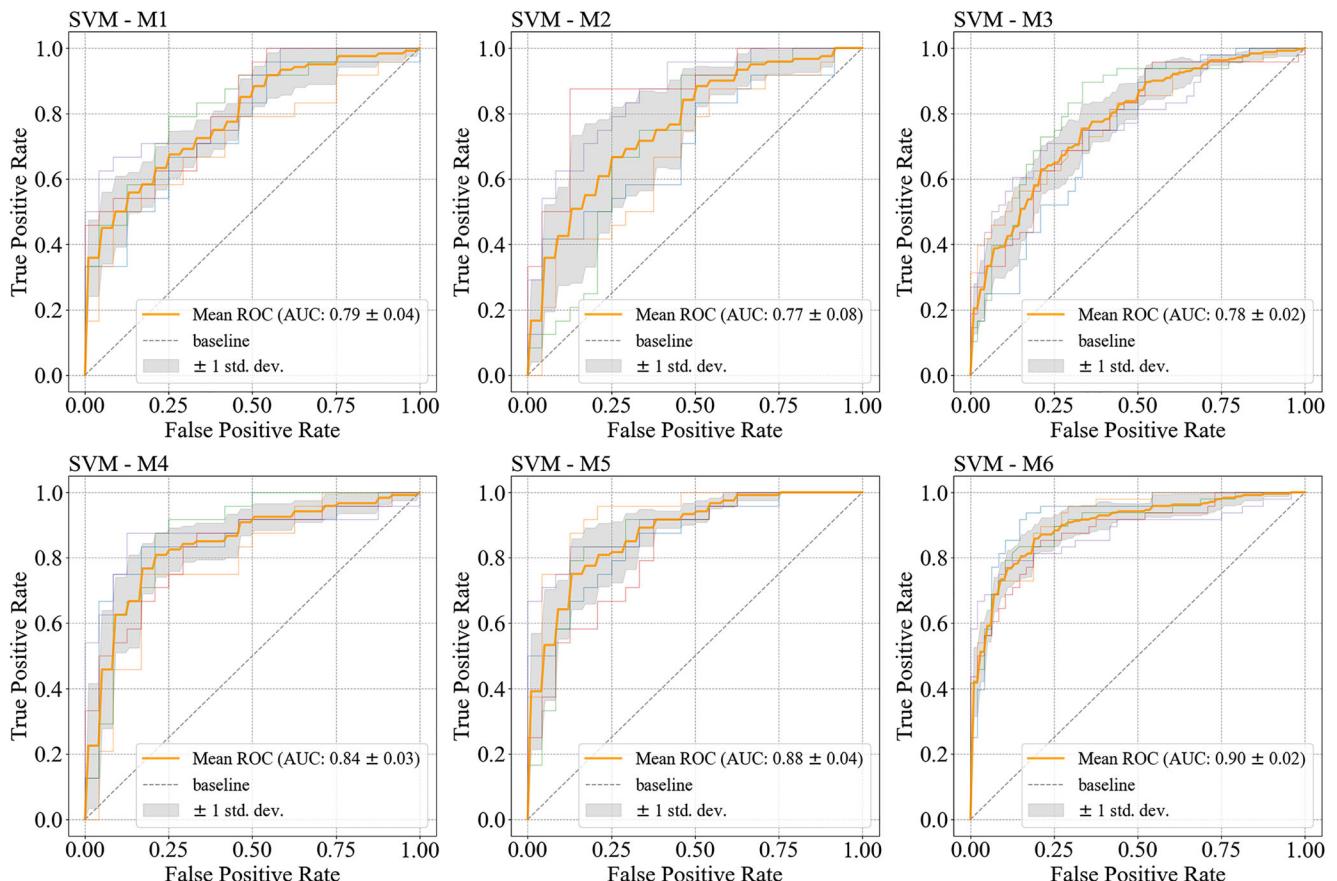


Fig. 8 ROC curve and AUC values of the SVM model. The M₁, M₂, and M₃ graphs represent the results of spiral drawing test set of the data collected from session 1, session 2, and all data together, respectively.

Similarly, M₄, M₅, and M₆ show the results of the sinusoidal drawings on the test set of the data collected from session 1, session 2, and all data together

Table 5 These results are from a test set of 1D CNN

Model	Session	Type of drawing	Accuracy (ACC)			Sensitivity (SEN)			Specificity (SPE)		
			Min	Max	Mean ± SD	Min	Max	Mean ± SD	Min	Max	Mean ± SD
M ₁	1	Spiral	0.687	0.813	0.750 ± 0.053	0.542	0.750	0.700 ± 0.081	0.625	0.875	0.800 ± 0.093
M ₂	2	Spiral	0.500	0.708	0.633 ± 0.078	0.417	0.958	0.830 ± 0.207	0.125	0.875	0.442 ± 0.256
M ₃	1, 2	Spiral	0.677	0.792	0.735 ± 0.042	0.688	0.813	0.746 ± 0.057	0.604	0.896	0.725 ± 0.124
M ₄	1	Sinusoidal	0.646	0.896	0.754 ± 0.083	0.292	0.917	0.580 ± 0.208	0.875	1.000	0.933 ± 0.057
M ₅	2	Sinusoidal	0.646	0.833	0.750 ± 0.069	0.500	1.000	0.717 ± 0.181	0.292	0.958	0.783 ± 0.249
M ₆	1, 2	Sinusoidal	0.802	0.875	0.831 ± 0.028	0.792	0.938	0.854 ± 0.056	0.729	0.875	0.808 ± 0.046

SD standard deviation

verifying the normal distribution, the homogeneous variance across groups, and the absence of outliers [49].

Table 7 shows the accuracy of all models that followed the requirements for the application of ANOVA. Therefore, ANOVA one-way was applied for each classifier, and for most of them, the *p* value was smaller than the significance level 0.05, leading to the conclusion that there are significant differences between the models. However, 2D CNN does not show significant differences between the accuracy means between models.

The Tukey honest significant difference method was employed to evaluate the differences between the accuracy of models. The results are presented in Fig. 11a and b for RFC, Fig. 11c and d for KNN, Fig. 11e and f for the SVM classifier, Fig. 11g and h for 1D CNN, and Fig. 11i and j for 2D CNN.

In Fig. 11, the boxplots show the accuracy values for each model that helps identify the differences in the Tukey graph. The label M₂ – M₁, for instance, denotes the average of model

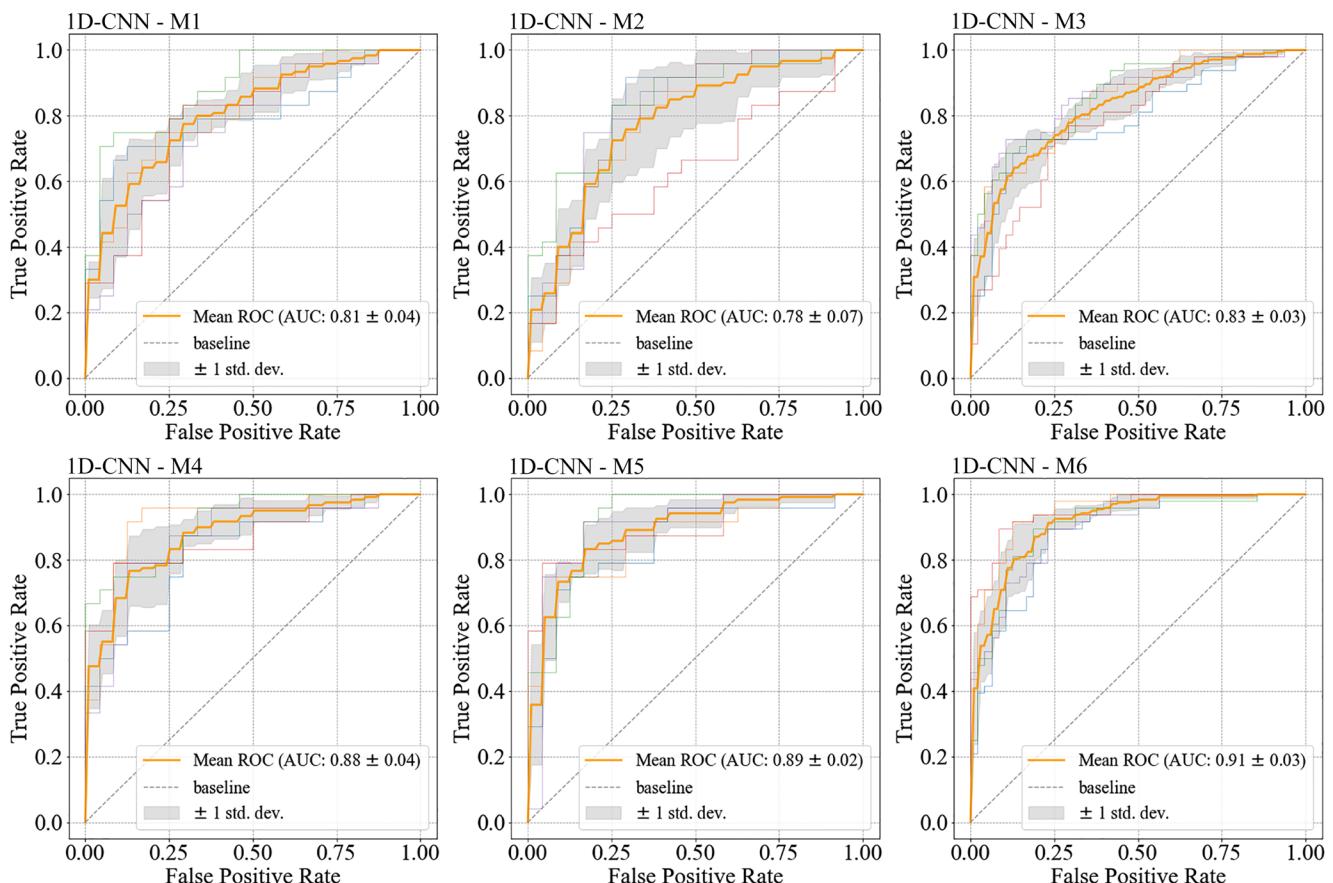


Fig. 9 ROC curve and AUC values of the 1D CNN model. The M₁, M₂, and M₃ graphs represent the results of spiral drawing test set of the data collected from session 1, session 2, and all data together, respectively.

Similarly, M₄, M₅, and M₆ show the results of the sinusoidal drawings on the test set of the data collected from session 1, session 2, and all data together

Table 6 These results are from a test set of 2D CNN

Model	Session	Type of drawing	Accuracy (ACC)			Sensitivity (SEN)			Specificity (SPE)		
			Min	Max	Mean ± SD	Min	Max	Mean ± SD	Min	Max	Mean ± SD
M ₁	1	Spiral	0.687	0.875	0.762 ± 0.063	0.625	0.875	0.775 ± 0.097	0.583	0.958	0.750 ± 0.137
M ₂	2	Spiral	0.687	0.812	0.754 ± 0.043	0.625	0.875	0.742 ± 0.113	0.667	0.917	0.766 ± 0.097
M ₃	1, 2	Spiral	0.729	0.813	0.763 ± 0.033	0.729	0.896	0.812 ± 0.066	0.604	0.771	0.713 ± 0.056
M ₄	1	Sinusoidal	0.733	0.854	0.804 ± 0.054	0.625	1.000	0.817 ± 0.136	0.708	0.875	0.792 ± 0.059
M ₅	2	Sinusoidal	0.708	0.833	0.767 ± 0.052	0.625	1.000	0.783 ± 0.127	0.583	0.917	0.750 ± 0.126
M ₆	1, 2	Sinusoidal	0.782	0.896	0.824 ± 0.039	0.771	0.896	0.835 ± 0.043	0.750	0.917	0.817 ± 0.056

SD standard deviation

2 minus model 1. Thus, if M₂ has a higher average than M₁, the difference is positive, otherwise negative.

4 Discussion

In this study, paper and pencil were used to quantify tremor through handwriting images. The HOG descriptor was

introduced as a novel feature for tremor detection from sinusoidal and spiral patterns. The features were extracted using HOG and classified by machine learning methods (i.e., RFC, KNN, SVM, and CNN) to detect tremor in people with PD. As the classification results based on sinusoidal patterns were larger than those based on spiral patterns, this research reinforces the necessity of looking for alternative drawings to the traditional spiral pattern.

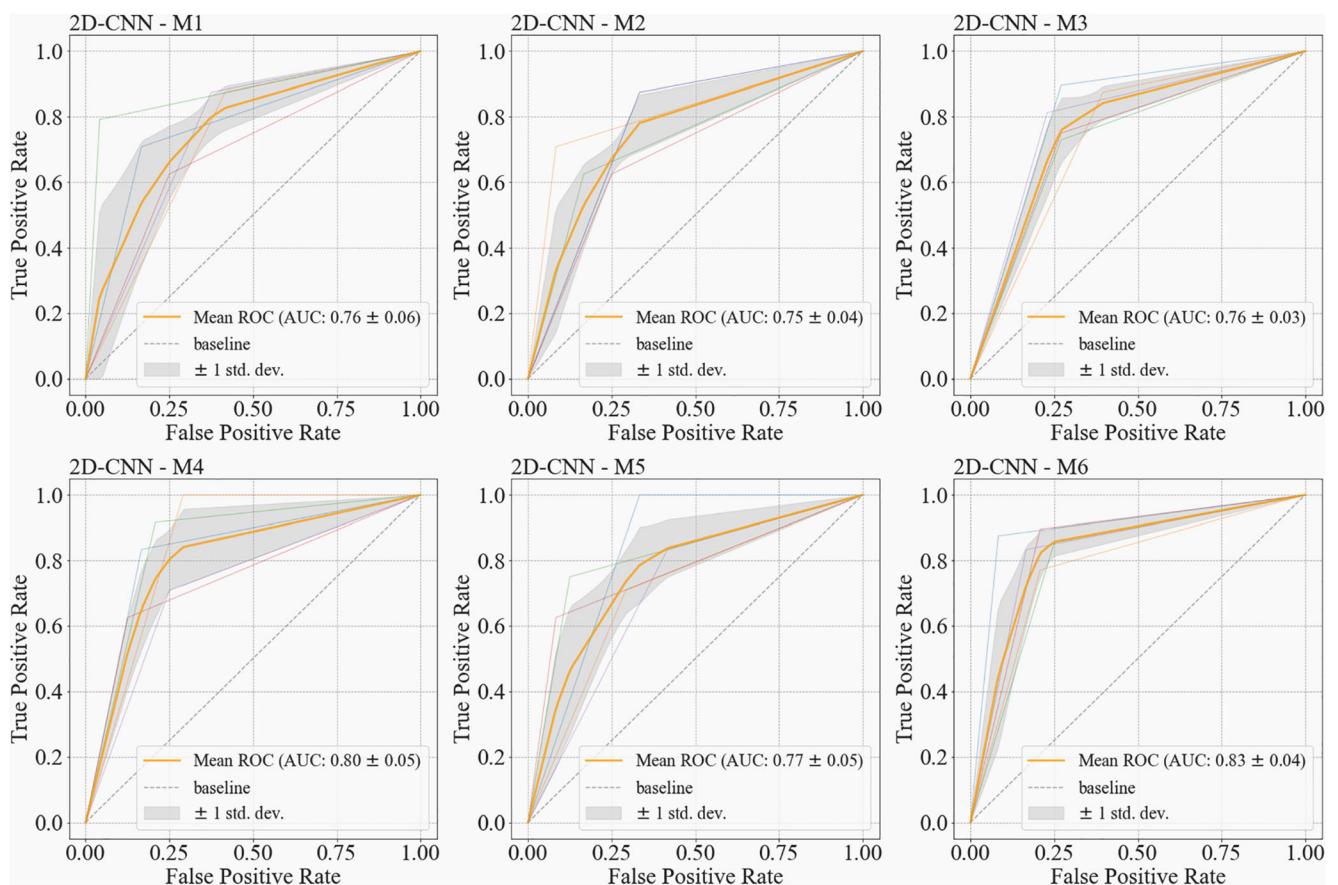


Fig. 10 ROC curve and AUC values of the 2D CNN model. The M₁, M₂, and M₃ graphs represent the results of spiral drawing test set of the data collected from session 1, session 2, and all data together, respectively.

Similarly, M₄, M₅, and M₆ show the results of the sinusoidal drawings on the test set of the data collected from session 1, session 2, and all data together

Table 7 Statistical tests to verify normality, variance, and presence of outliers

Classifier	Shapiro-Wilk		Bartlett's test		Bonferroni Outlier		ANOVA fit	ANOVA one-way	
	W	p value	B	p value	K	p value		F value	p value
RFC	0.973	0.630	5.988	0.307	2.038	1.000	Yes	3.585	0.015
KNN	0.949	0.161	3.032	0.695	2.860	0.265	Yes	4.537	0.005
SVM	0.970	0.530	10.042	0.074	3.382	0.077	Yes	4.213	0.007
1D CNN	0.987	0.967	5.335	0.376	2.548	0.539	Yes	4.233	0.007
2D CNN	0.963	0.369	1.911	0.861	2.601	0.479	Yes	1.609	0.196

W, B, and K are coefficient values of Shapiro-Wilk, Bartlett, and Bonferroni test. F value is the ANOVA result

The hand tremor of people with PD was confirmed by three distinct examiners through the application of the MDS-UPDRS, which is the gold standard method [10]. Such a clinical assessment is not available in several related studies that analyzed spiral patterns [9, 10, 12, 14–17, 19–23, 25], making it difficult for comparative evaluations. In this research, it was found a fair agreement between assessments from the evaluators as pointed out by Kendall's coefficient (66.1%), guaranteeing thus, the presence of a slight to a mild level of tremor in the participants with PD, as suggested by the mean sum of scores in MDS-UPDRS (session 1, 15.57; session 2, 16.42, values that vary between 0 and 48).

A practical and relevant aspect of this research is that there is no need for a supervised environment. Several studies [5, 6, 8–16] use tablet-based devices, which may require strict conditions to ensure the quality and usability of the acquired data. Other researches [12–14, 21, 22] employ less complex computing methods for data collection. However, they do not compare spiral patterns with additional drawings. In this sense, from the literature review, this is the first work to present objective results of handwritten drawings in comparison through the HOG descriptor.

Findings shown in Table 5 suggest that the result with model M₆ using the 1D CNN method yielded the best average values of ACC 83.1%, SEN 85.4%, and SPE 80.8%. These results are more accurate than those obtained from RFC (Table 2), KNN (Table 3), SVM (Table 4), and 2D CNN (Table 6). In addition, the highest true positive rates (91%) were obtained for RFC and 1D CNN, as suggested by the AUC estimates for M₆ in Figs. 6 and 9, respectively. Despite the small differences between the accuracies reached by the classifiers, the average time of training the models was elevated in some classifiers: 59.8 s for KNN, 105.7 s for RFC, 138.8 s for SVM, 1300.15 s for 1D CNN, and 16,957.88 s for 2D CNN in the environment described in this research.

Table 8 summarizes results reported by different studies, including those obtained from this research, which is the only one that compares results of sinusoidal and spiral drawings by using a unique type of visual feature (HOG). Conversely to distinct methods of feature extraction for tremor analysis [11, 12, 20, 23], HOG is independent of the type of the drawing

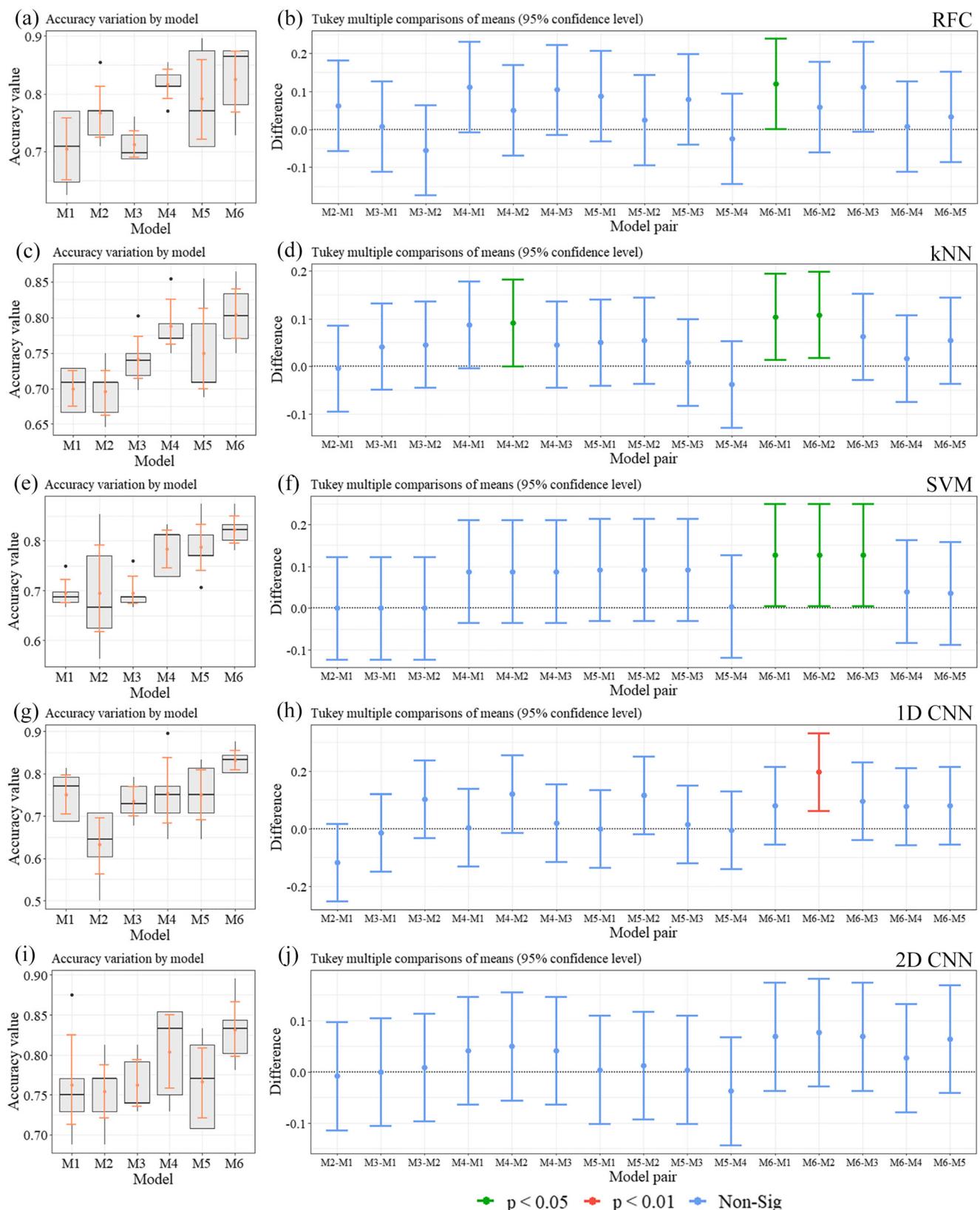
and does not need to be estimated online such as in [11, 12]. The use of devices for online estimate features, such as digitizing tablets, can limit the use of technology in the clinical scenario.

Two critical limitations of related researches are the lack of UPDRS evaluation of tremor and balanced groups. This research confirmed, through the assessment of three distinct examiners, that PwPD had slight to a mild level of tremor. It may be more challenging to discriminate these types of tremor from physiological tremor found in the control group. Furthermore, the lack of balanced groups may introduce bias in the results, thus, the comparison of outcomes from distinct studies is not straightforward, for instance, a study that used visual features estimated from digitized drawings reported accuracy results such as 83% [13] in balanced datasets, while 89.5% was reported [21] in a similar study that did not use balanced dataset. This research obtained, from a balanced data set, AUC of 83% for the spiral, as shown in Fig. 9 1D CNN-M3, and AUC of 91% for the sinusoidal, as illustrated in Fig. 9 1D CNN-M6. This suggests that the sinusoidal drawing should be considered in the clinical evaluation of patients.

Table 6 shows the results of CNN 2D; however, as shown in Fig. 10, the ROC curve and AUC indicate lower classification performance compared with the other tested classifiers employing HOG features. The fact that handwriting drawings have insufficient visual properties to be learnt from 2D CNN can justify this. The data set in this work only consists of 480 pictures per drawing (spiral or sinusoidal) and 240 images per experimental session with 120 pictures per group (PwPD and CG). This is a small dataset that can feed a 2D CNN which normally needs several thousand images per group [44].

Figure 4 shows important differences between features of PD and CG groups, and these differences were confirmed in Fig. 5 that shows the mean (a and b) and the coefficient of variation (c and d) for both groups and both types of handwritten drawings. The differences between groups are more considerable for features estimated from the sinusoidal than spiral images.

ANOVA was applied to evaluate the differences between the means of accuracies yielded by each model (Fig. 11). RFC ANOVA test and Tukey's analysis highlighted a mean



● $p < 0.05$ ● $p < 0.01$ ● Non-Sig

Fig. 11 Accuracy values of M₁–M₆ models for RFC (a and b), KNN (c and d), SVM (e and f), 1D CNN (g and h), and 2D CNN (i and j). The figure shows boxplots of the accuracy values with a vertical orange line that is

representing the mean and standard deviation. The Tukey multiple pairwise comparisons of the differences between means are also presented. Red and green colors highlight significant differences, whereas the blue non-significant

Table 8 Performance comparison of classifiers reported in different studies

Study	Drawing	Features	Classifier	Balanced groups	Results	UPDRS evaluation of hand tremor
[11]	Spiral	Online spatial, temporal, and pressure features using tablet-based	Logistic regression	No	AUC 83.7%, ACC -	Yes
[12]	Spiral	Online writing dynamics by pen and tablet-based	CNN	No	ACC 88%	No
[13]	Spiral	CNN features extracted from images produced from online coordinates	SVM	Yes	ACC 83%	No *
[20]	Spiral	Mean relative tremor and spatial features for spirals	Naïve Bayes	No	ACC 78.9%	No
[21]	Spiral, meander	Features extracted from digitized images	CNN	No	ACC meander 79.62%, ACC spiral 89.5%	No
[22]	Spiral, meander	Image features estimated from ResNet-50 (DNN)	OPF	No	ACC spiral 96.7%, ACC meander 96.3%	No *
[23]	Spiral	Fourier transform-based distance feature, tremor estimating distance-based feature	SVM	No	ACC 81.6%	No
This research	Sinusoidal, spiral	Histogram of oriented gradients (HOG)	1D CNN	Yes	ACC sinusoid 83.1%, ACC spiral 75%	Yes

The evaluation of results from sinusoidal handwritten drawings has only been reported in this research

CNN convolutional neural network, SVM support vector machine, DNN deep neural network, OPF optimum-path forest, AUC area under the curve, ACC accuracy

*It was applied the modified Hoehn and Yahr staging of severity

significant difference between M₆ (sinusoidal drawing model related to data collected in the sessions 1 and 2) and M₁ (spiral drawings model of session 1) in Fig. 11b.

The application of ANOVA to the results of KNN showed significant differences between means of the models M₆ and M₁, as shown in Fig. 11d; between M₆ and M₂ (the spiral drawing model related to data collected in session 2), and between M₄ (the sinusoidal drawing model related to data collected in session 1) and M₂. These results reinforce that the sinusoidal waveform can be more suitable than spiral drawings to evaluate the hand tremor symptom.

Results of SVM ANOVA showed more evidence of significant differences between M₆ and M₁, M₆ and M₂, and M₆ and M₃, as shown in Fig. 11f. Besides that, 1D CNN ANOVA also yielded a stronger significant difference between M₆ and M₂. All differences are related to spiral against sinusoidal-waveform subset models. ANOVA and Tukey tests could confirm similar significant differences between the means of spiral and sinusoidal drawing models. Despite the visual differences between the accuracy mean of the models, shown in Fig. 11j, the Tukey test applied to the 2D CNN models did not show any statistically significant difference, and this was confirmed in Table 7 that shows results of the ANOVA test with the null hypothesis accepted (p value > 0.05). On the other hand, the motor fluctuation could not be observed once there were no significant differences in the same type of drawing in distinct experimental sessions.

The literature reveals a lack of studies related to the significant differences between types of drawings. Smits et al. [5] used a tablet and pen to record the movement dynamics from circle,

stars, spirals, and letters, and they were able to distinguish between PD from CG. However, the authors did not have the objective to present a statistical comparison between the shapes. Pereira et al. [14, 21] also did not aim to present a statistical comparison between means of spirals and meanders. Passos et al. [22] described that they had not observed any significant difference in recognition rates between meanders and spirals.

Furthermore, in Fig. 11f, the spiral models presented the differences close to zero compared to each other (M₂ – M₁, M₃ – M₁, and M₃ – M₂), and the sinusoidal models presented their differences between means close to 0.05 (M₆ – M₄ and M₆ – M₅). New possible research could investigate if the SVM sinusoidal model is capable of assessing motor fluctuation between different experimental sessions of data collecting, which could help understanding of drug effects to patients with PD.

In the proposed experimental conditions, sinusoidal patterns are more appropriate for the detection of tremor. This may be explained by the fact that the sinusoidal drawing requires different skills from the participant when he has to slide the hand from one point to another on the paper sheet.

Another important point is that this approach does not require the attachment of sensors to the body of the participant. This may be relevant to the prevention of skin irritation, especially in the elderly, and also does not need equipment to perform the online evaluation.

Future work could be related to the development of a practical mobile application. The app can capture pictures from sinusoidal drawings and present quantitative tremor measurements to the clinician responsible for the follow-up of people

with PD. Due to the simplicity of the data collection procedure, this strategy could be used to monitor the patient at home without the need to care about controlling the equipment parameters such as speed, time, and pressure as needed in studies that use a tablet, digitizer pen, or inertial sensors.

5 Conclusion

The results of this research showed that HOG features extracted from sinusoidal handwriting drawings allow for better detection of tremor in people with PD. The highest results were obtained from the 1D CNN using HOG features (ACC 83.1%, SEN 85.4%, and SPE 80.8%, AUC 91%).

The methods described in this research were applied to a balanced dataset, and the feature extraction was based on a computer-vision technique, which was able to detect the tremor on the images. In addition, ANOVA and Tukey analysis evidenced the models from sinusoidal drawing obtained better results when compared with the models from spiral drawings, which is the most employed pattern in the clinical evaluation of tremor in people with PD.

All these results point that a sinusoidal pattern should be considered in the routine of clinical evaluations. Besides, the HOG descriptor can be applied with any drawing pattern other than the spiral and sinusoidal enhancing its use in assessing tremor in people with PD, unlike most studies that have been using specific features only for the drawings of Archimedean spiral.

Acknowledgments The present work was carried out with the support of the National Council for Scientific and Technological Development (CNPq), Coordination for the Improvement of Higher Education Personnel (CAPES—Program CAPES/DFATD-88887.159028/2017-00, Program CAPES/COFECUB-88881.370894/2019-01) and the Foundation for Research Support of the State of Minas Gerais (FAPEMIG-APQ-00942-17). A. O. Andrade, A. A. Pereira, and M. F. Vieira are fellows of CNPq, Brazil (304818/2018-6, 310911/2017-6, and 306205/2017-3).

Author contributions João Paulo Folador was responsible for the coding of programs for data analysis and storage, aiding in data collection, and writing and revising the paper. This work is one of the primary outcomes of his Ph.D. thesis.

Maria Cecilia Souza Santos, Luiza Maire David Luiz, and Luciane Aparecida Pascucci Sande de Souza are specialists in health sciences and Parkinson's disease. They were responsible for designing the experimental protocols, assessing patients through the application of the UPDRS, and data collection and paper revision.

Adriano Alves Pereira and Marcus Fraga Vieira contributed with the writing and revision of the paper.

Adriano de Oliveira Andrade was the principal supervisor of the study, contributing to all the steps from its conception to its conclusion, in addition to the paper writing and revision.

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflicts of interest.

Ethical approval The study was approved by the Research Ethics Committee of the Federal University of Uberlândia (CAAE: 07075413.6.0000.5152).

Informed consent Informed consent was obtained from all individual participants included in the study.

References

- Andrade AO, Pereira AA, de Almeida MFS, Cavalheiro GL, Paixão APS, Fenelon SB, Dionisio VC (2013) Human tremor: origins, detection and quantification. In: Practical applications in biomedical engineering. In Andrade AO, Pereira AA, Naves ELM, Soares AB (eds.) IntechOpen. <https://doi.org/10.5772/54524>
- Officer A, Wu D (2018) The global network for age-friendly cities and communities: looking back over the last decade, looking forward to the next. 34. (WHO/FWC/ALC/18.4). Licence: CC BY-NC-SA 3.0 IGO
- Gelb DJ, Oliver E, Gilman S (1999) Diagnostic criteria for Parkinson disease. *Arch Neurol* 56:33–39. <https://doi.org/10.1001/archneur.56.1.33>
- Navarro A, Castano Y, Valderrama J et al (2019) Objective levodopa response in Parkinson's disease: a study within the medical consultation using an RGB-D camera (Kinect®). In: International IEEE/EMBS Conference on Neural Engineering, NER. IEEE, pp 122–125
- Smits EJ, Tolonen AJ, Cluitmans L et al (2014) Standardized handwriting to assess bradykinesia, micrographia and tremor in Parkinson's disease. *PLoS One* 9. <https://doi.org/10.1371/journal.pone.0097614>
- Saunders-Pullman R, Derby C, Stanley K et al (2008) Validity of spiral analysis in early Parkinson's disease. *Mov Disord* 23:531–537. <https://doi.org/10.1002/mds.21874>
- Almeida MFS, Cavalheiro GL, Pereira AA, Andrade AO (2010) Investigation of age-related changes in physiological kinetic tremor. *Ann Biomed Eng* 38:3423–3439. <https://doi.org/10.1007/s10439-010-0098-z>
- Westin J, Ghiamati S, Memedi M et al (2010) A new computer method for assessing drawing impairment in Parkinson's disease. *J Neurosci Methods* 190:143–148. <https://doi.org/10.1016/j.jneumeth.2010.04.027>
- Surangsirat D, Thanawattano C (2012) Android application for spiral analysis in Parkinson's disease. In: Conference Proceedings - IEEE SOUTHEASTCON
- Surangsirat D, Intarapanich A, Thanawattano C et al (2013) Tremor assessment using spiral analysis in time-frequency domain. In: Conference Proceedings - IEEE SOUTHEASTCON
- San Luciano M, Wang C, Ortega RA et al (2016) Digitized spiral drawing: a possible biomarker for early Parkinson's disease. *PLoS One* 11. <https://doi.org/10.1371/journal.pone.0162799>
- Khatamino P, Canturk I, Ozylmaz L (2018) A deep learning-CNN based system for medical diagnosis: an application on Parkinson's disease handwriting drawings. 2018 6th Int Conf Control Eng Inf Technol 1–6. <https://doi.org/10.1109/CEIT.2018.8751879>
- 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 Recogn Lett* 121:19–27. <https://doi.org/10.1016/j.patrec.2018.04.008>
- Pereira CR, Weber SAT, Hook C et al (2017) Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics. In: Proceedings - 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI 2016. IEEE, pp 340–346

15. Tolonen A, Cluitmans L, Smits E et al (2015) Distinguishing Parkinson's disease from other syndromes causing tremor using automatic analysis of writing and drawing tasks. In: 2015 IEEE 15th International Conference on Bioinformatics and Bioengineering, BIBE 2015. IEEE, pp 1–4
16. Matsumoto Y, Tamura M, Fukumoto I (2009) Development of pathological evaluating system of tremor disease by a tablet PC and accelerometers. In: IFMBE Proceedings. pp 473–476
17. Kraus PH, Hoffmann A (2010) Spiralometry: computerized assessment of tremor amplitude on the basis of spiral drawing. *Mov Disord* 25:2164–2170. <https://doi.org/10.1002/mds.23193>
18. Bain PG, Findley LJ, Atchison P et al (1993) Assessing tremor severity. *J Neurol Neurosurg Psychiatry* 56:868–873. <https://doi.org/10.1136/jnnp.56.8.868>
19. Bajaj NPS, Knöbel M, Gontu V, Bain PG (2011) Can spiral analysis predict the FP-CIT SPECT scan result in tremulous patients? *Mov Disord* 26:699–704. <https://doi.org/10.1002/mds.23507>
20. Pereira CR, Pereira DR, Silva FAD et al (2015) A step towards the automated diagnosis of parkinson's disease: analyzing handwriting movements. In: Proceedings - IEEE Symposium on Computer-Based Medical Systems. Institute of Electrical and Electronics Engineers Inc., pp 171–176
21. Pereira CR, Pereira DR, Papa JP et al (2016) Convolutional neural networks applied for Parkinson's disease identification. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp 377–390
22. Passos LA, Pereira CR, Rezende ERS et al (2018) Parkinson disease identification using residual networks and optimum-path forest. In: SACI 2018 - IEEE 12th International Symposium on Applied Computational Intelligence and Informatics, Proceedings. Institute of Electrical and Electronics Engineers Inc., pp 325–329
23. Gupta J Das, Chanda B (2019) Novel features for diagnosis of Parkinson's disease from off-line Archimedean spiral images. In: 2019 IEEE 10th International Conference on Awareness Science and Technology, iCAST 2019 - Proceedings. Institute of Electrical and Electronics Engineers Inc
24. Daroff RB, Jankovic J, Mazziotta JC et al (2016) Bradley's neurology in clinical practice, 7th edn. Elsevier, London
25. Folador JP, Rosebrock A, Pereira AA et al (2020) Classification of handwritten drawings of people with Parkinson's disease by using histograms of oriented gradients and the random forest classifier. In: IFMBE Proceedings. pp 334–343
26. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005. IEEE, pp 886–893
27. Salfikar I, Sulistijono IA, Basuki A (2018) Automatic samples selection using histogram of oriented gradients (HOG) feature distance. *Emit Int J Eng Technol* 5:234–254. <https://doi.org/10.24003/emitter.v5i2.182>
28. Mokhtari M, Razzaghi P, Samavi S (2013) Texture classification using dominant gradient descriptor. In: Iranian Conference on Machine Vision and Image Processing, MVIP. IEEE, pp 100–104
29. Xiao Z, Yang Z, Geng L, Zhang F (2017) Traffic sign detection based on histograms of oriented gradients and Boolean convolutional neural networks. In: Proceedings - 2017 International Conference on Machine Vision and Information Technology, CMVIT 2017. Institute of Electrical and Electronics Engineers Inc., pp 111–115
30. Zhao Y, Chen D, Xie H et al (2019) Mammographic image classification system via active learning. *J Med Biol Eng* 39:569–582. <https://doi.org/10.1007/s40846-018-0437-3>
31. Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
32. Bernard S, Heutte L, Adam S (2009) Influence of hyperparameters on random forest accuracy. International Workshop on Multiple Classifier Systems (MCS), Jun 2009, Reykjavik, Iceland, pp 171–180. https://doi.org/10.1007/978-3-642-02326-2_18. (hal-00436358)
33. Ramani RG, Sivagami G (2011) Parkinson Disease classification using data mining algorithms. *Int J Comput Appl* 32(9):17–22. Available from: <https://www.ijcaonline.org/archives/volume32/number9/3932-5571>
34. Raczyk E, Zagajewski B (2017) Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *Eur J Remote Sens* 50:144–154. <https://doi.org/10.1080/22797254.2017.1299557>
35. Darnall ND, Donovan CK, Aktar S et al (2012) Application of machine learning and numerical analysis to classify tremor in patients affected with essential tremor or Parkinson's disease. *Gerontechnology* 10:208–219. <https://doi.org/10.4017/gt.2012.10.4.002.00>
36. Jeon H, Lee W, Park H et al (2017) High-accuracy automatic classification of Parkinsonian tremor severity using machine learning method. *Physiol Meas* 38:1980–1999. <https://doi.org/10.1088/1361-6579/aa8e1f>
37. Spyers-Ashby JM, Stokes MJ, Bain PG, Roberts SJ (1999) Classification of normal and pathological tremors using a multidimensional electromagnetic system. *Med Eng Phys* 21:713–723. [https://doi.org/10.1016/S1350-4533\(00\)0004-7](https://doi.org/10.1016/S1350-4533(00)0004-7)
38. Drotář P, Mekyska J, Rektorová I et al (2015) Decision support framework for Parkinson's disease based on novel handwriting markers. *IEEE Trans Neural Syst Rehabil Eng* 23:508–516. <https://doi.org/10.1109/TNSRE.2014.2359997>
39. Goetz CG, Fahn S, Martinez-Martin P et al (2007) Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): process, format, and clinimetric testing plan. *Mov Disord* 22:41–47. <https://doi.org/10.1002/mds.21198>
40. Kim TK (2015) T test as a parametric statistic. *Korean J Anesthesiol* 68:540–546. <https://doi.org/10.4097/kjae.2015.68.6.540>
41. Brodersen KH, Ong CS, Stephan KE, Buhmann JM (2010) The balanced accuracy and its posterior distribution. In: Proceedings - International Conference on Pattern Recognition. pp 3121–3124
42. Legendre P (2005) Species associations: the Kendall coefficient of concordance revisited. *J Agric Biol Environ Stat* 10:226–245. <https://doi.org/10.1198/108571105X46642>
43. Chowdhury SA, Kowsar MMS, Deb K (2018) Human detection utilizing adaptive background mixture models and improved histogram of oriented gradients. *ICT Express* 4:216–220. <https://doi.org/10.1016/j.icte.2017.11.016>
44. Chollet F (2017) Deep learning with Python. In: 1st ed. Manning Publications Co., USA
45. Mullin MD, Sukthankar R (2000) Complete cross-validation for nearest neighbor classifiers. In: ICML. pp 639–646. Available from: <https://www.cs.cmu.edu/~rahuls/pub/icml2000-rahuls.pdf>
46. Mucha J, Mekyska J, Galaz Z et al (2018) Identification and monitoring of Parkinson's disease dysgraphia based on fractional-order derivatives of online handwriting. *Appl Sci* 8:2566. <https://doi.org/10.3390/app8122566>
47. Prashanth R, Dutta Roy S, Mandal PK, Ghosh S (2016) High-accuracy detection of early Parkinson's disease through multimodal features and machine learning. *Int J Med Inform* 90:13–21. <https://doi.org/10.1016/j.ijmedinf.2016.03.001>
48. Zhu W, Zeng N, Wang N (2010) Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS® implementations. In: Northeast SAS Users Group 2010: Health Care and Life Sciences. pp 1–9

49. Quinn GP, Keough MJ (2002) Experimental design and data analysis for biologists. In: Cambridge University Press
50. Driscoll WC (1996) Robustness of the ANOVA and Tukey-Kramer statistical tests. *Comput Ind Eng* 31:265–268. [https://doi.org/10.1016/0360-8352\(96\)00127-1](https://doi.org/10.1016/0360-8352(96)00127-1)
51. Jankovic J (2005) Motor fluctuations and dyskinesias in Parkinson's disease: clinical manifestations. *Mov Disord* 20:S11–S16. <https://doi.org/10.1002/mds.20458>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

João Paulo Folador is a computer engineer with experience in data science. He is a Ph.D. candidate in Biomedical Engineering with a focus on health informatics and Parkinson's disease.

Maria Cecilia Souza Santos is a physiotherapist specialized in the assessment and rehabilitation of Parkinson's disease.

Luiza Maire David Luiz is a physiotherapist and Ph.D. student in Biomedical Engineering, whose research objective is on assessing Parkinson's disease, ergonomics, and biomechanics.

Dr. Luciane Aparecida Pascucci Sande de Souza is a professor specialized in motor control, evaluation, and treatment of neuromuscular disorders.

Dr. Marcus Fraga Vieira is a professor in Biomedical Engineering with expertise in biomechanics, mathematical modeling, physical education, and motor learning.

Dr. Adriano Alves Pereira is a professor in Biomedical Engineering, with expertise in biomechanics, health technology assessment, and signal analysis.

Dr. Adriano de Oliveira Andrade is a professor in Biomedical Engineering with experience in various health technological projects and with expertise in biomedical signal processing and modeling, assistive technology, rehabilitation engineering, neuroengineering, and Parkinson's disease.