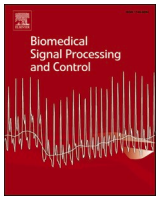
Biomedical Signal Processing and Control 75 (2022) 103551

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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



Tremor-related feature engineering for machine learning based Parkinson's disease diagnostics

Elli Valla [a,*], Sven Nõmm [a], Kadri Medijainen [b], Pille Taba [c], Aaro Toomela [d]

[a] Department of Software Science, School of Information Technology, Tallinn University of Technology (TalTech), Akadeemia tee 15a, 12618 Tallinn, Estonia
[b] Institute of Sport Sciences and Physiotherapy, University of Tartu, Ujula 4, Tartu 51014, Estonia
[c] Department of Neurology and Neurosurgery, University of Tartu, Puusepa 8, Tartu 51014, Estonia
[d] School of Natural Sciences and Health, Tallinn University, Narva mnt. 25, 10120 Tallinn, Estonia

ARTICLE INFO

Keywords:
Parkinson's disease
Handwriting database
Machine learning
Decision support system
Tremor

ABSTRACT

Growing research interest has arisen towards the possibility to automatically discriminate between the patients with neurodegenerative disease and healthy controls based on the information extracted from the digital drawing tests.

In this paper, we propose novel higher-order derivative based, angular-type and integral-like features extracted from the Archimedean spiral drawing tests for machine learning based Parkinson's disease diagnostics. The proposed features describe micro-changes in the handwriting trajectory, which are hard or impossible to detect with visual observation. However, they may hold valuable information in terms of tremor-like symptom analysis.

Two datasets are considered in this study: DraWritePD (acquired by the authors) and PaHaW (well known from the literature). A filter (Fisher's score) and wrapper (Recursive Feature Elimination) methods were used for feature selection. Six classifiers were trained and evaluated in a nested cross-validated loop to discriminate between healthy controls and Parkinson's patients.

A nested wrapper-type feature selection method combined with the ensemble classifiers predicted a disease with an accuracy of 84.33%, sensitivity of 70.00% and specificity of 93.20% (DraWritePD), and accuracy of 73.71%, sensitivity of 75.00% and specificity of 71.43% (PaHaW). The non-nested feature selection showed an over-optimistically high performance for both datasets: an accuracy of 92.16% (DraWritePD) and 84.86% (PaHaW).

The proposed novel tremor-related features were among the best performing predictors in the case of both datasets. Furthermore, the results indicate that the nested feature selection procedure plays a significant part in the classification performance.

1. Introduction

Neurodegenerative diseases, such as Alzheimer's and Parkinson's, are a class of neurological disorders where neurons from the central nervous system die or are damaged, causing severe disabilities [1]. Parkinson's disease (PD) is the second most common neurodegenerative disease after Alzheimer's. PD has a prevalence of approximately 0.5 to 1% among persons 65 to 69 years of age, rising to 1 to 3% among persons 80 years of age and older [2]. Several studies have produced evidence that pinpoints neurological disorders as one of the greatest burdens on the healthcare system. The Global Burden of Disease study suggests that 6.2 million patients are diagnosed with PD, and this number will double by 2040, surpassing the growth of Alzheimer's disease [3]. Finding accurate biomarkers for early diagnosis may significantly improve clinical intervention and treatment and can be utilised to monitor the progress of the disease [1,4]. Disorders in motor function performance, such as tremor, bradykinesia (slowness of movement), and rigidity (muscular stiffness), are the cardinal symptoms of PD [1,5]. Tremor occurs in approximately 75% of patients with PD [5–7]. According to the study, [5], early diagnosed PD patients find tremor the second most troublesome symptom, and advanced PD patients (progressed over six years) ranked it as the first motor-related symptom that diminishes the quality

* Corresponding author.
E-mail addresses: elli.valla@taltech.ee (E. Valla), sven.nomm@taltech.ee (S. Nõmm), kadri.medijainen@ut.ee (K. Medijainen), pille.taba@kliinikum.ee (P. Taba), aaro.toomela@tlu.ee (A. Toomela).

https://doi.org/10.1016/j.bspc.2022.103551
Received 18 October 2021; Received in revised form 11 January 2022; Accepted 2 February 2022
Available online 13 February 2022

of their life. Traditionally, medical diagnosis is based on subjective observations from different clinical tests. For example, standardised handwriting tasks can provide quantitative measures for the assessment of tremor [8]. In their classical pen and paper setting, these pure human assessments suffer from several drawbacks: the presence of a subjective component; the limits of human perception, like inability to measure velocity, pressure applied, not to mention derivatives such as jerk, shake, etc.; completion time being the only precisely measurable parameter of the test; and finally, there is no clear definition nor description of errors in these type of assessments. It has been well documented that the digital signals extracted from the handwriting of PD patients are affected and therefore might serve as a diagnostic marker in a computer-aided analysis [9–11]. While existing results may lead highly accurate results [16,19,21] not a lot attention has been paid to the feature selection process, making achieved results less interpretable and attractive for the medical community. While in some papers, features are selected purely based on their discriminating power, many other contributions omit in-depth discussion of the feature selection process. In this paper, we also compare non-nested and nested feature selection processes to confirm the importance of the chosen methods in a classification pipeline (Section 3.3). Based on the digital Archimedean spiral drawing test [22], we demonstrate that the proposed novel tremor-related features possess high discriminating power and provide accurate diagnostics support.

The rest of this paper is organised as follows. Section 1.1 reviews the works related to this problem. Sections 2 and 3 respectively describe the materials and methods used in this research. Sections 4 and 5 report and discuss the experimental results to highlight the effectiveness of the proposed method.

1.1. State of the art

Drawing and writing tests have been used in psychology and neurology for at least a century [23]. The analysis of these handwritten tasks has proven effective in the diagnosis and progression monitoring of PD patients [13,14]. Spiral drawing tests have been frequently used for studying motor control deficits in PD patients [9,12]. It is proved to be a useful tool for assessing tremor-related symptoms [24]. Dynamic methods make use of digital tablets [18], smartpens with axial pressure of ink and tri-axial accelerometers [25], tablet computers [19] and other devices [9]. The main advantage of online acquisition devices is their ability to acquire dynamics of the writing process, which are lost with offline systems. More specifically, dynamic features are the position of the pen (coordinates), pressure (force applied on the writing surface), azimuth (pen orientation), altitude (pen inclination), and timestamp [10]. The most commonly used approaches in the relevant studies can be categorised as follows: numeric methods, where kinematics of the handwriting are analysed [16–19] and deep learning based approach [21,25], where the image or time-series data is extracted and used for classification. In numerical analysis, a typical processing chain involves the pre-processing of raw signals followed by feature extraction and classification. Most proposed methods make use of supervised learning techniques, such as Logistic Regression, AdaBoost, Naive Bayes, SVM, KNN, Random Forests, Decision Trees, LDA. By far the most used classifier in all researched diseases is SVM [9,10,12]. Deep neural networks are not as popular as the more conventional algorithms mentioned before [9,10], but they are starting to gain their popularity [21,25–27].

One of the first contributions describing the results of the digitised drawing test was Marquardt et al. (1994) [15]. More than 30 years of research has resulted in the fact that the original set of four parameters proposed by [15] has grown significantly to hundreds of parameters [16,18,19].

Significant research has been done by Drotar et al. [18], where the various features for the prediction of PD are extracted using digital tablets. Kinematic and pressure parameters were computed from in-air as well as on-surface time intervals. The classification was carried out by applying Support Vector Machine (SVM), and the maximum accuracy of the spiral drawing test was 62.9%. The authors also suggested that classification accuracy depends on the choice of the template. Ensemble of all tests obtained an accuracy of 81.3% on the kinematic and pressure features. The study was conducted on the PaHaW database with a sample size of 75 test subjects.

In the research of Nõmm et al. [28,19] the set of features initially proposed in [29] to measure the quantity and smoothness of the human motions observed during the gross motor activity were adapted for the case of fine motor motions observed during drawing tests. The main distinctive component of [28,19] is the tuple integral-like parameters referred as *motion mass*. On the example of Luria's alternating series tests, it was demonstrated by [19] that motion mass parameters possess higher discriminating power compared to commonly used average values and time duration of the test.

Impedovo (2019) [16] investigates a wide set of velocity-based features for PD patients and healthy controls (HC) discrimination. The extended feature set includes parameters obtained from the Sigma-Lognormal model, the Maxwell–Boltzmann distribution, and the Discrete Fourier Transform applied to the velocity profile of handwriting. The prediction was 97.3% accurate based on the spiral drawing test.

Rios-Urrego et al. (2019) [30] proposed to use geometrical and non-linear dynamic features in addition to kinematic features. It was assumed that these features are able to capture the irregularities of handwriting, which increase as the disease advances. The results showed that the kinematic features were most accurate and that it is possible to discriminate between Parkinson's patients and healthy controls with accuracies up to 83.3%.

Angelillo et al. (2019) [17] investigated the predictive potential of an optimal subset of tasks for an automatised PD diagnosis. First, several features exploiting the dynamics of the handwriting process are extracted from the raw data of different tasks. Then, the predictive potential of each task is evaluated individually. Finally, the best tasks, i.e., those with the highest prediction accuracy, are fed into an ensemble of classifiers whose predictions are obtained via majority voting. Experiments were performed on the PaHaW dataset, as it includes several tasks performed by the same subjects. Non-nested and nested cross-validation performance were compared and analysed. Overall performance degradation was noticeable and therefore concluded the importance of the nested feature selection step in the cross-validation. Poor performance was obtained by the spiral task. The accuracy score of 53.75% was achieved using SVM with RBF kernel, confirming the findings already reported in [18]. Assembling all tasks for classification showed significant improvement in performance. The best performing model with SVM (linear kernel) achieved 88.75% accuracy. It is important to note that it was obtained with non-nested feature selection, as was done in [18,16]. Ensemble of tasks performance with nested feature selection was 66.25% and was obtained with SVM_{RBF}.

Yang et al. (2019) [31] extracted features using polar expressions from the hand-drawn Archimedean spiral and straight-line drawing tests. Features including deviation (cm), accumulation angle (rad), and drawing velocity (cm/s), were engineered and differentiated for normal control groups and groups with Parkinson's disease or essential tremor. The SVM-based classifier showed promising results, with a mean true negative rate (specificity) of 86.79% for identifying normal controls, a mean true positive rate (sensitivity) of 93.72% for identifying PD and ET cases, and a mean hit rate of 90.84% for identifying the correct classes. In [20] hash transformation is used to map polar expression features to a high-dimensional space. The proposed decision-making classifier achieved a higher mean true negative rate (98.96%), mean true positive rate (98.93%), and mean hit rate (98.93%). The important thing to note here is that the difference between the mean age of control subjects and Parkinson's patients who participated in the study is 15.27 (17.57 in [31]) years, and there is no overlap in terms of standard deviation. This is a quite significant difference considering that the performance of

human motor functions declines with age [32–34]. High classification accuracy in their studies may be partly related to the fact that the fine motor skills of Parkinson's patients were impaired both because of the disease and because of age; this fact might have increased the differences between the groups.

For the sake of fair comparison, we included the studies that used statistical machine learning methods, excluding the papers with deep learning algorithms. Other criteria for inclusion are the use of the Archimedean spiral task from the PaHaW dataset and a clear description of the feature selection procedure. For this reason, results of the following studies were chosen for comparison: [16–18].

2. Materials

Two datasets are considered in this research. The first one, here and after referred to as *DraWritePD*, was acquired by the authors. The second dataset, *PaHaW*, is well known from the literature and was kindly provided by the authors of [35,18]. The PaHaW [18] database was used as an additional dataset to test the stability and performance of the proposed method.

2.1. Drawing and handwriting tests for Parkinson's diagnostics, DraWritePD

Data acquisition was performed with an Apple iPad Pro (2016) tablet computer and an Apple Pencil. The tablet has a 26.77 cm (10.5 inches) diagonal. The iPad Pro scans the Apple Pencil's signal with a frequency of 240 points per second. From a software perspective - data was collected using a custom iOS application developed by the research team. The dynamic features (time-sequences) captured by the tablet are as follows: x-coordinate (mm); y-coordinate (mm); timestamp (sec); pressure (arbitrary unit of force applied on the surface: [0,…, 6.0]); altitude (rad); azimuth (rad). Total of 24 PD patients (mean age 74.1 $\pm$ 6.7) and 34 age- and gender-matched healthy control subjects (mean age 74.1 $\pm$ 9.1)) participated in the creation of the database. The overall task was to complete a testing battery consisting of 12 different drawing and writing tests. In this paper, we focus only on the Archimedean spiral test. Few sample images of healthy and Parkinson's patient's spiral drawings are depicted in Fig. 1. The data acquisition process was conducted with the strict guidance of privacy law. The Research Ethics Committee of the University of Tartu (No. 1275T −9) approved the study.

2.2. Parkinson's disease handwriting (PaHaW) database

Data acquisition of the *PaHaW* dataset is described in detail in [18,35]. For the sake of self-sufficiency main properties of this dataset important for the present studies are described in this section. Age and gender distribution of the *PaHaW* dataset is similar to those of *DraWritePD*. The data set consists of 37 PD patients and 38 age- and gender-matched healthy controls (HC). HC subjects have a mean age of 62.4 years (standard deviation 11.3), whereas PD patients have a mean age of 69.3 years (standard deviation 10.9) [18]. During the acquisition of *PaHaW* dataset, each subject was asked to complete a handwriting task according to the prepared pre-filled template at a comfortable speed. Subjects were allowed to repeat the task in case of some error or mistake during handwriting [18]. The handwriting signals were recorded using a Wacom Intuos digitising tablet overlaid with a blank sheet of paper, the sampling rate was set to 100 samples per second. The tablet captured the following dynamic features: x-coordinate; y-coordinate; timestamp; button status; pressure; altitude, and azimuth. All features were converted to the same units as in DraWritePD. The battery of the tasks presented in *PaHaW* dataset differs much from the one employed in *DraWritePD*. However, the Archimedean spiral drawing test is present in both datasets and was thus used in this research.

3. Methods

The methodology used in the present study consists of three main stages: feature engineering, feature selection, and classification. Details of each stage are presented in the following subsections and depicted in Fig. 2.

3.1. Feature engineering

Raw time-series described in Section 2.1: pen position (x- and y-coordinates), timestamp, pen pressure, pen inclination (altitude), and pen orientation (azimuth) can be used to compute an infinite number of features. Tables 1, 2 represent feature classes selected for the present research. Kinematic (displacement, velocity, acceleration, etc.), spatial–temporal (duration, distance), geometric (altitude, azimuth, yaw, etc.), and pressure features were derived. Feature extraction resulted in either a single-valued feature or a vector feature. For all resulting vector features, the following statistical measures were calculated: mean, median, standard deviation, maximum and minimum value. In addition, horizontal and vertical components of the kinematic features were computed. In this research, each subject was instructed to draw a spiral in one stroke; therefore, stroke-related and on/off-screen time values, that are explored in [35,18], are omitted.

Our contribution to this feature engineering is the addition of the higher-order derivatives with respect to time. For instance, given a respective timestamp we can calculate the velocity of the position vector $\overrightarrow{r} = [p_i, p_{i+1}]$. In other words, velocity is the rate at which displacement of the position vector changes with respect to time. Similarly, acceleration was computed as the rate of change in velocity and jerk as the rate of change in acceleration with respect to time. This approach is visualised in Fig. 3. Following the sequence, we considered up to the sixth time derivative of the position vector. There are no universally accepted names for the fourth and higher time derivatives of the displacement. However, the terms snap, crackle, and pop are used in literature for the fourth, fifth, and sixth time derivatives of displacement [37]. The same

(a) HC (b) PD (c) PD

Fig. 1. Sample drawings of an Archimedean spiral performed by a healthy control subject (a) and the PD patients (b, c) from DraWritePD dataset.

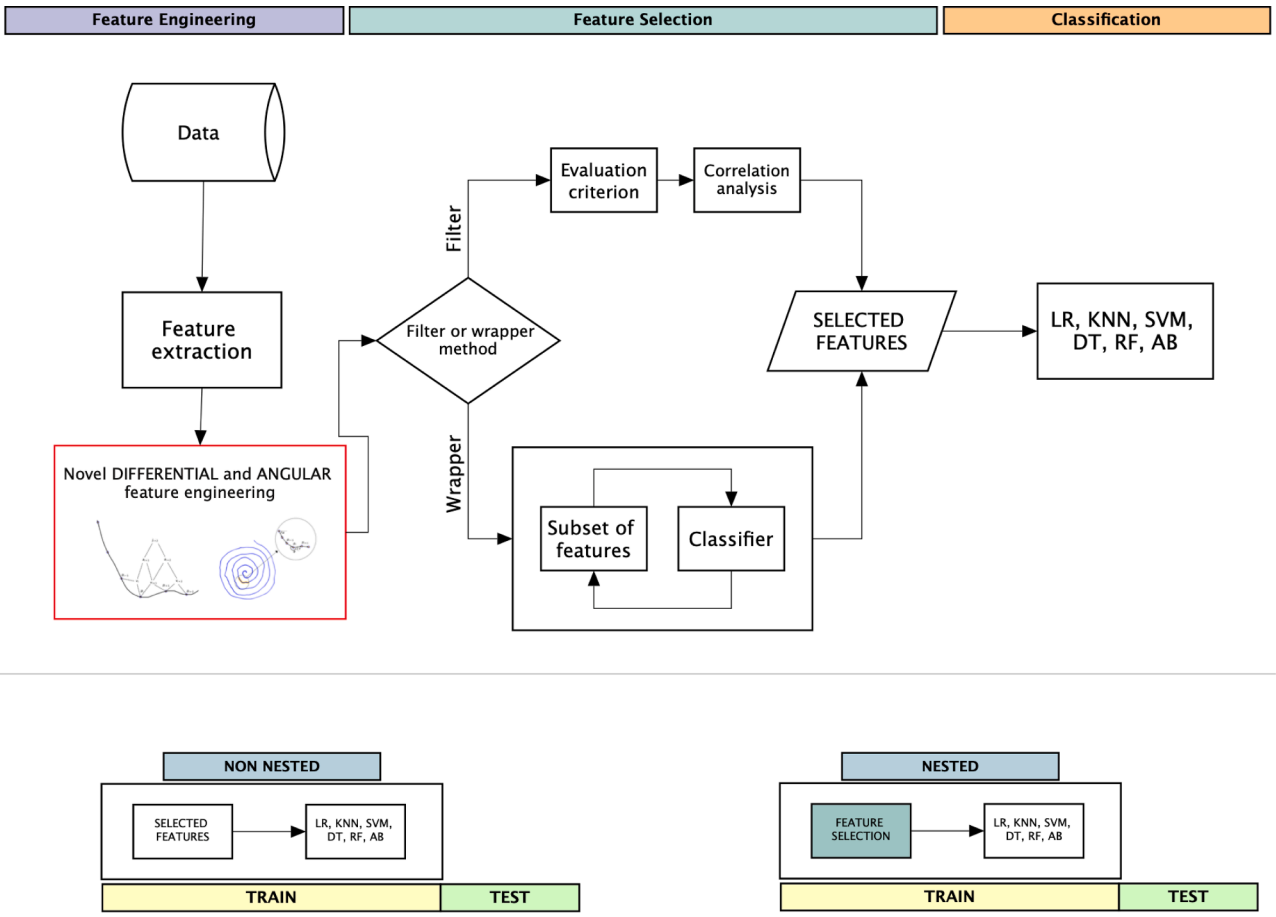


Fig. 2. General methodology. The novel features engineered are described in Section 3.1. Non-nested and nested feature selection visualisation depicts the difference between the two approaches. In the case of the non-nested method, the features are selected based on all the samples and then used for training. The correct way (nested) would be to perform feature selection inside the cross-validation fold using only the samples from the train set, keeping the test set "unseen". [36].

approach was taken for calculating additional pressure parameters. Up to fourth-time derivatives of pressure (force applied on the surface) were computed. These derivatives are referred to as yank, tug, snatch, and shake, respectively [37].

It is worth noting that pen inclination and orientation angles are not widely exploited in related studies. In addition to altitude (pen inclination) and azimuth (pen orientation), we are considering three angles of the position vectors. Pen's trajectory can be observed in Fig. 3 (note that in Fig. 3 magnified part was downsampled for illustration purposes). Given the slope k of the position vector, we can extract angle α. Let N be the number of observation points and (x_i, y_i) are the coordinates of the point p_i, where $i \in \{1, 2, \ldots, N\}$, then the slope (k) and respective angle are represented as follows:

$$k = \frac{y_i - y_{i-1}}{x_i - x_{i-1}}, \tag{1}$$

$$\alpha = \arctan k, \tag{2}$$

Fig. 3 depicts other two angles that are considered in the current research: a rotational angle ϕ (3) and a yaw angle γ (4):

$$\phi_i = \pi + \alpha_{i-1} - \alpha_i \tag{3}$$

$$\gamma_i = \alpha_i - \alpha_{i-1} \tag{4}$$

Yaw is described as the change in direction in which the point vector is pointing. The angular feature set was also enriched with up to third respective time derivatives.

These micro-changes in changes or differential features alongside the angular features of the drawing trajectory are hard or even impossible to detect with the visual observation; however, they may hold valuable information in terms of tremor-like symptom analysis. Tremor associated with the PD is usually reflected by the less smooth motions, which in turn means greater accelerations, more directional changes, and increases in other similar parameters. This symptom is often described as an involuntary quivering, shaky movement and rhythmical declination of trajectory, which gives the basis to relate proposed differential and angular features with the disease. The results indicate that proposed features can be successfully exploited in addition to kinematic and pressure information to enrich feature representation further.

A study conducted by [19] showed that the tuple of integral-like features computed based on kinematic parameters and pressure possess sufficiently high discriminating power to distinguish PD patients from HC control subjects. In [38] it was also demonstrated that these features might allow machine learning techniques to detect mental fatigue; therefore, these features are included in the present research. For the sake of self-sufficiency, the computational procedure of the motion parameters is described in the following subsection.

3.1.1. *Motion mass parameters*

Motion mass parameters were introduced by [29] to describe the amount and smoothness of motion of a limb or some other group of joints. For each kinematic, geometric, and pressure parameter that changes during the test sum of the absolute values at each observation point may be computed. Let N be the number of observation points in the test (or a part of the test). Denote v_k the velocity along the directional vector of the stylus movement at observation point k where $k \in \{1, \ldots, N\}$

Table 1
Sample subset of vector features.

Feature set	Feature	Description
Spatial-temporal features	displacement	$d_i = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}$
Kinematic features	velocity	Rate of change in displacement with respect to time. First time derivative of the displacement.
	acceleration	Rate of change in velocity with respect to time. Second time derivative of the displacement.
	jerk	Rate of change in acceleration with respect to time. Third time derivative of the displacement.
	snap	Rate of change in jerk with respect to time. Fourth time derivative of the displacement.
	crackle	Rate of change in snap with respect to time. Fifth time derivative of the displacement.
	pop	Rate of change in crackle with respect to time. Sixth time derivative of the displacement.
Pressure features	pressure_diff	Change in pressure between points $[p_i, p_{i+1}]$
	yank	Rate of change in pressure. First time derivative of the force applied on the surface.
	tug	Rate of change in yank. Second time derivative of the force applied on the surface.
	snatch	Rate of change in tug. Third time derivative of the force applied on the surface.
	shake	Rate of change in snatch. Fourth time derivative of the force applied on the surface.
Geometric features	altitude_diff	Change in altitude angle between points $[p_i, p_{i+1}]$
	azimuth_diff	Change in the azimuth angle between points $[p_i, p_{i+1}]$
	alphas_diff	Change in alpha angle between points $[p_i, p_{i+1}]$
	phi_angle_diff	Change in phi angle between points $[p_i, p_{i+1}]$
	yaw_diff	Change in yaw angle between points $[p_i, p_{i+1}]$

then *velocity mass* is defined by equation

$$V_N = \sum k = 1^N |v_k| \tag{5}$$

In the same way, mass parameters are defined for the acceleration - A_N, jerk - J_N, yank Y_N, tug - T_N, snatch - Sn_N and shake - Sh_N. The same logic applies to the pressure (force applied on the surface) and angular parameters describing changes in the stylus direction. In [19] trajectory length and time duration were combined with velocity, acceleration, jerk, and pressure masses into the tuple referred as *motion mass*. The present research adds mass parameters for the higher derivatives of the accelerations and treats them as the features to be used by machine learning methods.

Table 2
Sample subset of single-value features.

Feature set	Feature	Description
Spatial-temporal features	duration	Time interval between first and last time stamp signal
Kinematic features	velocity_mass	Velocity mass of the point vector $[p_1, p_2, \ldots, p_k, \ldots, p_N]$
	acceleration_x_mass	Mass of the x-directional rate of change in velocity
	jerk_median	Median value of the rate of change in acceleration
	snap_mass	Mass of the fifth time derivative of displacement.
Pressure features	shake_median	Median value of the fifth time derivative of pressure (force applied on the surface).
	pressure_diff_min	Minimum difference of pressure between points $[p_i, p_j]$
	tug_mass	Mass of the second time derivative of pressure (force applied on the surface)
Geometric features	ϕ_mass	Mass of the angle ϕ (in radians), see Fig. 3
	α_accel_min	Minimum acceleration of the angle α, see Fig. 3
	yaw_std	Standard deviation of the yaw, see Fig. 3

3.2. Feature selection

A total of 202 features were extracted from the raw signals. Most discriminative predictors were selected to reduce dimensionality and increase interpretability using filter- and wrapper-type feature selection procedures.

Filter methods use independent statistical techniques to evaluate the relationship between a feature predictor and a target variable. They are widely used among bioinformatics researchers due to their straightforward and computationally inexpensive implementation [39]. Among filter models suggested by [40], the Fisher score is the most natural choice due to the numerical representation of the features. For each feature, the Fisher score is computed by the following Eq. (6). Large values of the Fisher's score indicate higher discriminating power and, therefore, better suitability for machine learning classifiers. The algorithm used in this paper returns the ranks of the variables based on Fisher's score in descending order.

$$F = \frac{\sum_{j=1}^{k} p_j (\mu_j - \mu)^2}{\sum_{j=1}^{k} p_j \sigma_j^2} \tag{6}$$

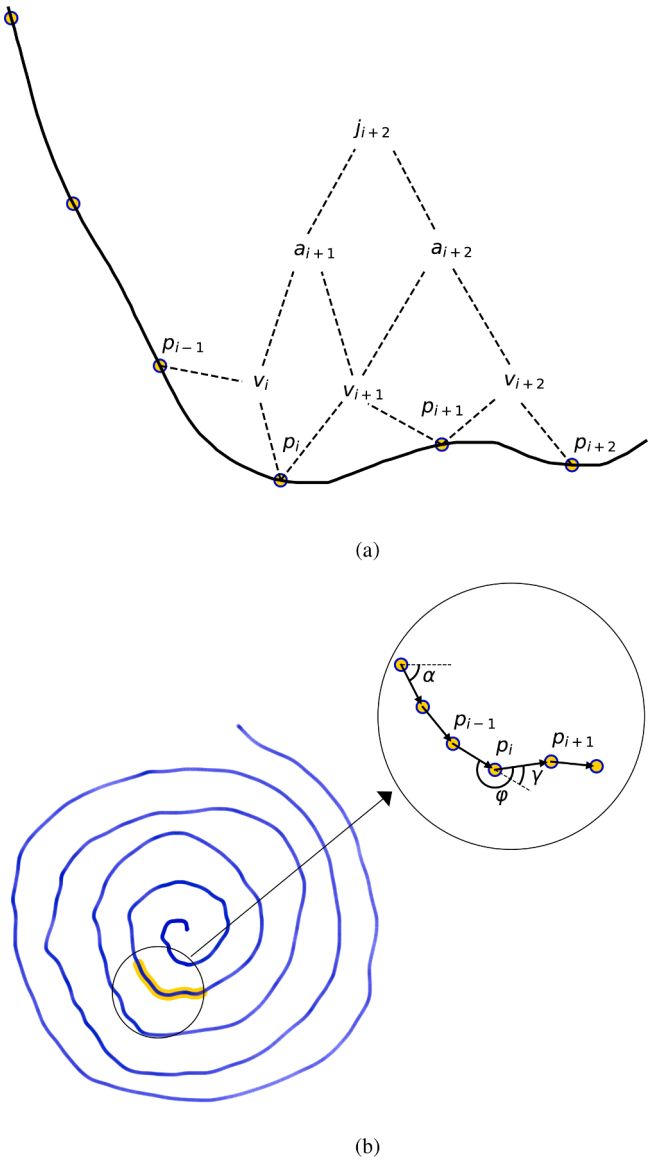


Fig. 3. Visual representation of the differential-type (a) and angular-type (b) features.

Pair-wise Pearson correlations among the sorted features were then computed and explored. A set of features with a Pearson correlation below the threshold of 0.7 was constructed.

Although filter methods are fast and scalable, they have the disadvantage of ignoring the interaction with the classifier. For this reason, wrapper methods are commonly utilised in the feature selection process. In this case, the evaluation is done by training and testing a specific classification model that estimates the relevance of a given subset of features. Wrapper methods are proven to be more efficient but also more computationally expensive [41]. In this paper, we consider the SVM recursive feature elimination (SVM-RFE) wrapper method proposed by [41].

3.3. Statistical classification

In the proposed framework, supervised feature selection strategies are 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 only for validation. The problem with the a priori or so-called non-nested feature selection is that the predictors have an unfair advantage, as they were chosen on the basis of all of the samples. This procedure does not correctly mimic the application of the classifier to a completely independent test set since these predictors "leak the information" from train to test set. [17,36] In other words, a non-nested feature selection may introduce a bias that may lead to overfitting a model and, therefore, to an over-optimistic performance. In this paper, we report results for both nested and non-nested feature selection to analyse this effect. Six machine learning classifiers Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), AdaBoost (AB), were trained and cross-validated in a k-fold loop, $k \in [3,5,10]$. Training and validation of the classifiers were performed using the scikit-learn library [42] for Python.

The performance of the classifiers was reported according to the following metrics: accuracy, precision, sensitivity, and specificity. Accuracy alone does not provide insights into the rate of true positive and true negative predictions by ignoring per-class performance evaluation. Statistical measures: specificity (true negative rate) - refers to the ability of the test to identify subjects without the disease correctly; sensitivity (recall or true positive rate) - refers to the ability of the test to identify those patients with the disease correctly; are therefore widely used in medical diagnostic settings.

4. Results

In this section, we report the results of a series of experiments aimed at evaluating the performance of the proposed workflow.

4.1. Numerical results

In the following Tables 3, 4, the mean accuracy, precision, sensitivity, and specificity values are reported, averaged over all the iterations of a k-fold cross-validation scheme. Feature set remained the same from one fold of cross-validation to another. Low variability is an indication of a stable feature selection algorithm.

The related state-of-the-art results obtained on the PaHaW dataset are depicted in Table 5.

4.2. Discussion

The degradation of the overall performance in the case of the nested feature selection indicates that the model built from the pre-selected features may have been overoptimistic. In addition, in Tables 3 and 4 it can be seen that the wrapper method outperformed the filter-type feature selection algorithm. Proposed tremor-related features were present for all feature selection procedures. For the DraWritePD dataset, the differential features shake, snap, crackle, and pop with their respective *motion mass* parameters selected by a nested wrapper-type feature selection combined with an ensemble classifier, demonstrated a high performance with an accuracy score of 84.33%, sensitivity 70.00%, and specificity of 93.33%. The same methodology performed on the PaHaW dataset showed an accuracy of 73.71%, a sensitivity of 75.00%, and specificity of 71.43%; the selected predictors contained a combination of the differential and angular features with their respective statistical measurements. Although Logistic Regression was among the best performing classifiers in a majority of the settings, in the case of nested wrapper-type feature selection, ensemble algorithms (AdaBoost and Random Forest) showed greater performance.

It is important to note, that Impedovo (2019) [16] and Drotar et al. (2016) [18] used the non-nested validation scheme. This means that, while the results reported by [16] (see Table 5) are remarkable, they still suffer feature selection bias described in Section 3.3.

The nearly perfect classification performance presented in the literature [16,20,21,26,43] is another topic for discussion. Clinical observations show that early-stage patients with Parkinson's disease under medication do not necessarily differ significantly from healthy

Table 3
Classification performance with non-nested and nested feature selection for the DraWritePD dataset. The best scores for each feature selection method are presented in bold.

		Features	Classifier	P_{acc}	P_{prec}	P_{sen}	P_{spec}
Filter method	non-nested	ϕ_mass, duration, pressure_median	LR	84.18%	88.33%	75.00%	90.00%
			RF	**88.18%**	**92.00%**	**80.00%**	**93.81%**
			KNN	78.36%	78.33%	65.00%	87.14%
			SVM	80.36%	88.33%	65.00%	87.14%
			DT	80.18%	72.00%	75.00%	83.81%
			AB	78.36%	71.00%	70.00%	83.81%
	nested	slopes_min, altitude_median, duration	LR	**68.63%**	**55.56%**	34.13%	**90.00%**
			RF	56.86%	33.99%	**61.90%**	56.67%
			KNN	49.02%	25.10%	42.86%	56.67%
			SVM	52.94%	27.06%	44.44%	57.58%
			DT	54.90%	13.73%	33.33%	66.67%
			AB	54.90%	13.73%	33.33%	66.67%
Wrapper method	non-nested	velocity_median, α_accel_max, pressure_median	LR	90.20%	**95.24%**	80.16%	**96.67%**
			RF	**92.16%**	91.67%	**90.48%**	93.94%
			KNN	80.39%	78.57%	80.16%	81.52%
			SVM	88.24%	90.48%	80.16%	93.94%
			DT	80.39%	76.72%	79.37%	81.52%
			AB	86.27%	86.11%	80.95%	90.61%
	nested	shake_mass, shake_max, snap_mass, crackle_mass, pop_mass	LR	55.33%	25.67%	45.00%	61.67%
			RF	80.33%	80.00%	65.00%	90.00%
			KNN	82.00%	81.67%	70.00%	90.00%
			SVM	84.00%	**86.67%**	**75.00%**	90.00%
			DT	72.67%	65.83%	55.00%	83.33%
			AB	**84.33%**	81.67%	70.00%	**93.33%**

Table 4
Classification performance with non-nested and nested feature selection for the PaHaW dataset. The best scores for each feature selection method are presented in bold.

		Features	**Classifier**	P_{acc}	P_{prec}	P_{sen}	P_{spec}
Filter method	non-nested	shake_mean, accel_x_mass, α_accel_max, shake_median, α_accel_min	LR	**79.17%**	**80.67%**	**77.78%**	**80.56%**
			RF	69.44%	72.50%	66.67%	72.22%
			KNN	61.11%	58.63%	75.00%	47.22%
			SVM	63.89%	62.16%	80.56%	47.22%
			DT	62.50%	62.63%	66.67%	58.33%
			AB	66.67%	67.64%	69.44%	63.89%
	nested	pressure_diff_min, shake_mean, shake_median	LR	**65.28**%	55.00%	67.50%	**65.00%**
			RF	58.33%	**61.64%**	63.89%	52.78%
			KNN	56.94%	42.59%	52.78%	61.11%
			SVM	50.00%	33.33%	66.67%	33.33%
			DT	55.56%	58.28%	**75.00%**	36.11%
			AB	52.78%	55.93%	69.44%	36.11%
Wrapper method	non-nested	accel_x_min, α_accel_min, pressure_diff_max, shake_mean, shake_max	LR	**84.86**%	**90.00%**	**80.36%**	**88.57%**
			RF	59.81%	62.61%	52.50%	66.43%
			KNN	68.29%	72.00%	55.00%	80.36%
			SVM	73.62%	78.44%	71.79%	74.29%
			DT	61.24%	63.79%	52.50%	69.64%
			AB	62.86%	63.33%	61.43%	64.29%
	nested	α_velocity_max, α_accel_min, snatch_mean	LR	65.33%	67.33%	60.71%	69.29%
			RF	**73.71%**	**76.62%**	**75.00%**	71.43%
			KNN	63.90%	63.33%	66.79%	60.71%
			SVM	66.76%	70.67%	60.71%	**72.14%**
			DT	66.86%	70.33%	60.36%	71.29%
			AB	64.10%	66.90%	58.57%	69.29%

Table 5
Performance comparison with the state-of-the-art methods based on the Archimedean spiral test from the PaHaW dataset.

	Drotar et al. (2016)	**Impedovo (2019)**	**Angellilo *et al.* (2019)**	**Present work**
non-nested	62.8	97.3	51.3	**84.9**
nested	–	–	53.8	**73.7**

individuals of comparable age and may perform even better than the fine motor skills of elderly individuals without Parkinson's disease. In other words, there is no clear-cut categorical difference between fine-motor skills of individuals with and without Parkinson's disease. Future work would involve testing methodologies introduced in these studies on our own (DraWritePD) dataset to investigate this conclusion further.

5. Conclusion

The handwriting- and drawing-based computer-aided analysis have the potential to serve as the decision support tool for clinicians in neurodegenerative disease diagnostics. Its successful implementation would play a significant role in reducing the burden on the public health system. Our study proposes a set of tremor-related features to discriminate Parkinson's disease patients (PD) and healthy controls (HC) based on the Archimedean spiral drawing test. More specifically, the set of variables was enriched with angular, differential, and integral-like parameters resulting in a database with over 200 features. To reduce the dimensionality and maximise the model's performance, we applied filter- and wrapper-type feature selection algorithms. As indicated in Table 3, the proposed tremor-related features were among the best performing predictors in the PD and HC classification task. It was also demonstrated that a non-nested feature selection method might lead to over-optimistic results and, therefore, should be avoided in the classification pipeline. These findings were reproduced on the PaHaW dataset, see Table 4. The classification accuracy obtained in the present research is 20% higher compared to those reported in the literature (case of a nested feature selection), which confirms that the novel tremor-related features possess greater discriminating power for the diagnostics of Parkinson's disease.

A couple of remarks have to be recognised when interpreting the results. The proposed framework was conducted based only on the Archimedean spiral test. In our future work, we plan to include other handwriting and drawing tasks. Repeated measurements from the same subjects should be obtained and analysed to confirm the possible relation of the proposed features with tremor-based motor dysfunctions and assess the severity of the symptom. To demonstrate the reliability and generalisability of the proposed framework, we need to perform testing on more extensive and more diverse groups of patients. A small sample size is a significant limitation of the present study. In this particular problem domain, developing a large labelled dataset for automatic machine learning based analysis is an ongoing issue. Our future work strives to overcome this obstacle by providing additional databases and new methods for data enhancement and augmentation. Despite these limitations, the reported performance values are indeed very promising. Concept-wise and from the soft- and hardware perspective, the proposed framework is ready for clinical use.

CRediT authorship contribution statement

Elli Valla: Software, Formal analysis, Investigation, Visualization, Writing - original draft, Writing - review & editing. **Sven Nõmm:** Methodology, Validation, Writing - review & editing. **Kadri Medijainen:** Data curation. **Pille Taba:** Supervision. **Aaro Toomela:** Conceptualization, Supervision.

Declaration of Competing Interest

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

Acknowledgments

This work in the project "ICT programme" was supported by the European Union through European Social Fund. The authors would also like to thank the BDALab team for generously providing the PaHaW dataset to help with our research endeavours.

References

[1] L.M. De Lau, M.M. Breteler, Epidemiology of parkinson's disease, Lancet Neurol. 5 (6) (2006) 525–535.
[2] R.L. Nussbaum, C.E. Ellis, Alzheimer's disease and parkinson's disease, New England J. Med. 348 (14) (2003) 1356–1364.
[3] E. Dorsey, A. Elbaz, E. Nichols, F. Abd-Allah, A. Abdelalim, J. Adsuar, M. Ansha, C. Brayne, J. Choi, D. Collado-Mateo, et al., Gbd 2016 parkinson's disease collaborators. global, regional, and national burden of parkinson's disease, 1990–2016: a systematic analysis for the global burden of disease study 2016, Lancet Neurol. 17 (11) (2018) 939–953.
[4] R. He, X. Yan, J. Guo, Q. Xu, B. Tang, Q. Sun, Recent advances in biomarkers for parkinson's disease, Front. Aging Neurosci. 10 (2018) 305.
[5] M. Politis, K. Wu, S. Molloy, P.G. Bain, K.R. Chaudhuri, P. Piccini, Parkinson's disease symptoms: the patient's perspective, Mov. Disord. 25 (11) (2010) 1646–1651.
[6] A. Gironell, B. Pascual-Sedano, I. Aracil, J. Marín-Lahoz, J. Pagonabarraga, J. Kulisevsky, Tremor types in parkinson disease: a descriptive study using a new classification, Parkinson's Disease 2018 (2018).
[7] A. Barbeau, et al., Parkinson's disease: clinical features and etiopathology, Handbook of clinical neurology 5 (49) (1986) 87–152.
[8] E. Smits, A. Tolonen, L. Cluitmans, M. Gils, B. Conway, R.C Zietsma, K. Leenders, N. Maurits, Standardized Handwriting to Assess Bradykinesia, Micrographia and Tremor in Parkinson's disease, PloS one 9 (05 2014). doi:10.1371/journal.pone.0097614.
[9] C. De Stefano, F. Fontanella, D. Impedovo, G. Pirlo, A.S. di Freca, Handwriting analysis to support neurodegenerative diseases diagnosis: A review, Pattern Recogn. Lett. 121 (2019) 37–45.
[10] D. Impedovo, G. Pirlo, Dynamic handwriting analysis for the assessment of neurodegenerative diseases: a pattern recognition perspective, IEEE Rev. Biomed. Eng. 12 (2018) 209–220.
[11] S. Rosenblum, M. Samuel, S. Zlotnik, I. Erikh, I. Schlesinger, Handwriting as an objective tool for parkinson's disease diagnosis, J. Neurol. 260 (9) (2013) 2357–2361.
[12] G. Vessio, Dynamic handwriting analysis for neurodegenerative disease assessment: A literary review, Appl. Sci. 9 (21) (2019) 4666.
[13] T. Eichhorn, T. Gasser, N. Mai, C. Marquardt, G. Arnold, J. Schwarz, W. Oertel, Computational analysis of open loop handwriting movements in parkinson's disease: a rapid method to detect dopamimetic effects, Movement Disorders 11 (3) (1996) 289–297.
[14] J. Phillips, G.E. Stelmach, N. Teasdale, What can indices of handwriting quality tell us about parkinsonian handwriting? Hum. Mov. Sci. 10 (2–3) (1991) 301–314.
[15] C. Marquardt, N. Mai, A computational procedure for movement analysis in handwriting, J. Neurosci. Methods 52 (1) (1994) 39–45, https://doi.org/10.1016/0165-0270(94)90053-1.
[16] D. Impedovo, Velocity-based signal features for the assessment of parkinsonian handwriting, IEEE Signal Process. Lett. 26 (4) (2019) 632–636.
[17] M.T. Angelillo, D. Impedovo, G. Pirlo, G. Vessio, Performance-driven handwriting task selection for parkinson's disease classification, in: International Conference of the Italian Association for Artificial Intelligence, Springer, 2019, pp. 281–293.
[18] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smékal, M. Faundez-Zanuy, Evaluation of handwriting kinematics and pressure for differential diagnosis of parkinson's disease, Artif. Intell. Med. 67 (2016) 39–46, https://doi.org/10.1016/j.artmed.2016.01.004.
[19] S. Nõmm, K. Bardõš, A. Toomela, K. Medijainen, P. Taba, Detailed analysis of the luria's alternating seriestests for parkinson's disease diagnostics, in: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018, pp. 1347–1352, https://doi.org/10.1109/ICMLA.2018.00219.
[20] T.-L. Yang, C.-H. Lin, W.-L. Chen, H.-Y. Lin, C.-S. Su, C.-K. Liang, Hash transformation and machine learning-based decision-making classifier improved the accuracy rate of automated parkinson's disease screening, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (1) (2019) 72–82.
[21] M. Diaz, M. Moetesum, I. Siddiqi, G. Vessio, Sequence-based dynamic handwriting analysis for parkinson's disease detection with one-dimensional convolutions and bigrus, Expert Syst. Appl. 168 (2021), 114405.
[22] J. Westin, S. Ghiamati, M. Memedi, D. Nyholm, A. Johansson, M. Dougherty, T. Groth, A new computer method for assessing drawing impairment in parkinson's disease, J. Neurosci. Methods 190 (1) (2010) 143–148, https://doi.org/10.1016/j.jneumeth.2010.04.027.
[23] E. Hazan, F. Frankenburg, M. Brenkel, K. Shulman, The test of time: a history of clock drawing, Int. J. Geriatric Psychiatry 33 (1) (2018) e22–e30.
[24] S.L. Pullman, Spiral analysis: a new technique for measuring tremor with a digitizing tablet, Mov. Disord. 13 (S3) (1998) 85–89.
[25] C.R. Pereira, S.A. Weber, C. Hook, G.H. Rosa, J.P. Papa, Deep learning-aided parkinson's disease diagnosis from handwritten dynamics, in: 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI) Ieee, 2016, pp. 340–346.
[26] I. Kamran, S. Naz, I. Razzak, M. Imran, Handwriting dynamics assessment using deep neural network for early identification of parkinson's disease, Future Gener. Comput. Syst. 117 (2021) 234–244.

[27] S. Zarembo, S. Nõmm, K. Medijainen, P. Taba, A. Toomela, Cnn based analysis of the luria's alternating series test for parkinson's disease diagnostics, in: Asian Conference on Intelligent Information and Database Systems Springer, 2021, pp. 3–13.
[28] S. Nõmm, A. Toomela, J. Kozhenkina, T. Toomsoo, Quantitative analysis in the digital luria's alternating series tests, in: 2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV), 2016, pp. 1–6, https://doi.org/10.1109/ICARCV.2016.7838746.
[29] S. Nõmm, A. Toomela, An alternative approach to measure quantity and smoothness of the human limb motions, Est. J. Eng. 19 (4) (2013) 298–308.
[30] C.D. Rios-Urrego, J.C. Vásquez-Correa, J.F. Vargas-Bonilla, E. Nöth, F. Lopera, J. R. Orozco-Arroyave, Analysis and evaluation of handwriting in patients with parkinson's disease using kinematic, geometrical, and non-linear features, Comput. Methods Programs Biomed. 173 (2019) 43–52.
[31] T.-L. Yang, P.-J. Kan, C.-H. Lin, H.-Y. Lin, W.-L. Chen, H.-T. Yau, Using polar expression features and nonlinear machine learning classifier for automated parkinson's disease screening, IEEE Sens. J. 20 (1) (2019) 501–514.
[32] N.S. Frolov, E.N. Pitsik, V.A. Maksimenko, V.V. Grubov, A.R. Kiselev, Z. Wang, A. E. Hramov, Age-related slowing down in the motor initiation in elderly adults, Plos one 15 (9) (2020), e0233942.
[33] T. Stöckel, K. Wunsch, C.M. Hughes, Age-related decline in anticipatory motor planning and its relation to cognitive and motor skill proficiency, Front. Aging Neurosci. 9 (2017) 283.
[34] Y.Y. Hoogendam, F. van der Lijn, M.W. Vernooij, A. Hofman, W.J. Niessen, A. van der Lugt, M.A. Ikram, J.N. van der Geest, Older age relates to worsening of fine motor skills: a population-based study of middle-aged and elderly persons, Front. Aging Neurosci. 6 (2014) 259.
[35] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, M. Faundez-Zanuy, Analysis of in-air movement in handwriting: A novel marker for parkinson's disease, Comput. Methods Programs Biomed. 117 (3) (2014) 405–411.
[36] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning. 2nd Edition, Springer Series in Statistics, Springer, 2002.
[37] R.N. Jazar, Advanced Dynamics. Rigid Body, Multibody, and Aerospace Applications, John Wiley & Sons Inc, 2007.
[38] O. Senkiv, S. Nõmm, A. Toomela, Applicability of spiral drawing test for mental fatigue modelling, IFAC-PapersOnLine 51 (34) (2019) 190–195, 2nd IFAC Conference on Cyber-Physical and Human Systems CPHS 2018. doi: 10.1016/j.ifacol.2019.01.064.http://www.sciencedirect.com/science/article/pii/S2405896319300679.
[39] K. Tadist, S. Najah, N.S. Nikolov, F. Mrabti, A. Zahi, Feature selection methods and genomic big data: a systematic review, J. Big Data 6 (1) (2019) 1–24.
[40] C. Aggarwal, Data Mining, Springer, 2015.
[41] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Mach. Learn. 46 (1) (2002) 389–422.
[42] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[43] L.S. Bernardo, A. Quezada, R. Munoz, F.M. Maia, C.R. Pereira, W. Wu, V.H.C. de Albuquerque, Handwritten pattern recognition for early parkinson's disease diagnosis, Pattern Recogn. Lett. 125 (2019) 78–84.