



ELSEVIER

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

A new computer vision-based approach to aid the diagnosis of Parkinson's disease

Clayton R. Pereira ^a, Danilo R. Pereira ^b, Francisco A. Silva ^b,
João P. Masieiro ^b, Silke A.T. Weber ^c, Christian Hook ^d, João P. Papa ^{e,*}

^a Department of Computing, Federal University of São Carlos, Brazil

^b University of Western São Paulo, Brazil

^c Department of Ophthalmology and Otorhinolaryngology, São Paulo State University, Brazil

^d Ostbayerische Technische Hochschule, Germany

^e Department of Computing, São Paulo State University, Brazil

ARTICLE INFO

Article history:

Received 2 March 2016

Received in revised form

9 June 2016

Accepted 11 August 2016

Keywords:

Parkinson's disease

Pattern recognition

Micrography

ABSTRACT

Background and Objective: Even today, pointing out an exam that can diagnose a patient with Parkinson's disease (PD) accurately enough is not an easy task. Although a number of techniques have been used in search for a more precise method, detecting such illness and measuring its level of severity early enough to postpone its side effects are not straightforward. In this work, after reviewing a considerable number of works, we conclude that only a few techniques address the problem of PD recognition by means of micrography using computer vision techniques. Therefore, we consider the problem of aiding automatic PD diagnosis by means of spirals and meanders filled out in forms, which are then compared with the template for feature extraction.

Methods: In our work, both the template and the drawings are identified and separated automatically using image processing techniques, thus needing no user intervention. Since we have no registered images, the idea is to obtain a suitable representation of both template and drawings using the very same approach for all images in a fast and accurate approach.

Results: The results have shown that we can obtain very reasonable recognition rates (around ≈67%), with the most accurate class being the one represented by the patients, which outnumbered the control individuals in the proposed dataset.

Conclusions: The proposed approach seemed to be suitable for aiding in automatic PD diagnosis by means of computer vision and machine learning techniques. Also, meander images play an important role, leading to higher accuracies than spiral images. We also observed that the main problem in detecting PD is the patients in the early stages, who can draw near-perfect objects, which are very similar to the ones made by control patients.

© 2016 Elsevier Ireland Ltd. All rights reserved.

* Corresponding author. Department of Computing, São Paulo State University, Brazil. Fax: +55-14-3103-6079.
E-mail address: papa@fc.unesp.br (J.P. Papa).

<http://dx.doi.org/10.1016/j.cmpb.2016.08.005>

0169-2607/© 2016 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Parkinson's disease (PD) is a degenerative, chronic, and progressive illness that may cause tremors, slowness of movement, muscle stiffness, and changes in speech and writing skills due to the neurological disorder [1]. PD was first described by the English physician James Parkinson [2], with its symptoms being well-known in the scientific community. However, to diagnose Parkinson's disease with a reliable recognition rate in its early stages is still unheard of. Moreover, it is not straightforward to establish the PD level soon after its diagnosis.

Parkinson's disease occurs when nerve cells that produce dopamine are destroyed, a process that is performed slowly, thus characterizing the progression of this disease. With the absence of such a substance, the nerve cells can no longer send messages properly, causing many other symptoms such as depression, sleep disturbances, memory impairment and autonomic nervous system disorders. In some cases, Parkinson's disease may be triggered by hereditary causes [1].

In the last decades, some works attempted at designing solutions to aid PD diagnosis. Expert systems based on machine learning techniques have been employed to this purpose, showing promising results [3]. Generally, these works are signal analysis-oriented, which means one can use the patient's voice to assess the level of the illness [4,5], since the voice capability is gradually compromised by PD. Little et al. [4], for instance, presented a dataset composed of biomedical voice measurements from 31 male and female subjects, of which 23 patients were diagnosed with PD and 8 were healthy subjects. The authors introduced a new measure of dysphonia called Pitch Period Entropy, which seems to be more robust in identifying changes in the speech, since approximately 90% of PD patients exhibit some form of vocal impairment [6,7].

In the work conducted by Zhao et al. [8], five patients and seven healthy individuals were used to recognize Parkinson's disease by means of voice analysis. In order to fulfill this purpose, voices of the patients were recorded using an Isomax EarSet E60P5L microphone; the recording sessions lasted around 25 minutes each, and the authors used a total of 50 pre-recorded prompts consisting of emotional sentences spoken by a professional actress. Tsanas et al. [9] evaluated different algorithms based on dysphonia measures aiming at PD recognition. A total of 132 acoustic features were initially used for further feature selection, and the authors concluded that the dysphonia information and the existing features end up helping PD recognition. Harel et al. [10] claimed that PD symptoms are detectable up to five years prior to clinical diagnosis, and symptoms presented in speech include reduced loudness, increased vocal tremor, and breathiness. In their work, the authors used a dataset of the National Center for Voice and Speech, which comprises 263 phonations from 43 subjects (17 females and 26 males, of which 10 were healthy controls and 33 were diagnosed with PD).

Since one of the first manifestation of Parkinson's Disease is the deterioration of handwriting, the micrography (a writing exam) is another approach widely used for the diagnosis of Parkinson's disease [11]. This technique is considered an objective measure, since a PD patient possibly features the reduction of calligraphy size, as well as the hand tremors. Nowadays, this

procedure is often conducted by filling out some specific forms. Rosenblum et al. [12] suggested that writing exams can be used to distinguish PD patients from healthy individuals. The authors employed the following methodology to support their assumption: 20 PD patients and 20 control individuals were asked to write their names and addresses in a piece of paper attached to a digital table. Further, for each stroke, the mean pressure and velocity were measured in order to compute spatial and temporal information. The authors presented very good recognition rates, with 97.5% of the participants classified correctly (100% of the control individuals, and 95% of PD patients). Later on, Drotár et al. [13] claimed that movement during handwriting of a text consists not only from the on-surface movements of the hand, but also from the in-air trajectories performed when the hand moves in the air from one stroke to the next. The authors demonstrated the assessment of in-air hand movements during sentence handwriting has a higher impact than the pure evaluation of on surface movements, leading to classification accuracies of 84% and 78%, respectively.

Machine learning-based techniques have also been applied to help automatic PD recognition. Spadotto et al. [14], for instance, introduced the Optimum-Path Forest (OPF) [15,16] classifier to the aforementioned context. Later on, Spadotto et al. [17] proposed an evolutionary-based approach to select the most discriminative set of features in order to improve PD recognition rates. Gharehchopogh and Mohammadi [18] used Artificial Neural Networks with Multi-Layer Perceptron to diagnose the effects caused by Parkinson's disease. Pan et al. [19] analyzed the performance of Support Vector Machines with Radial Basis Function in order to compare the onset of tremor in patients with Parkinson's disease. Hariharan et al. [20] developed a new feature weighting method using Model-based clustering (Gaussian mixture model) in order to enrich the discriminative ability of the dysphonia-based features, thus achieving 100% of classification accuracy. Recently, Peker et al. [21] used sound-based features and complex-valued neural networks to aid PD diagnosis as well.

However, although many works deal with voice- and speech-driven information, there is a large number of writing exams out there that can give us valuable information about the development of Parkinson's Disease, since it is cheaper and easier to acquire such sort of exam. Moreover, most hospitals and clinics have writing exams by hand only, which means they need to be digitized prior to information extraction. Usually, the patients are asked to draw spirals and meanders, which are then compared against the templates. Very recently, Pereira et al. [22] proposed to extract features from writing exams using image processing techniques, achieving around 79% of recognition rates, which is considered very reasonable. The authors also designed and made available a dataset called "HandPD" with all images and features extracted.¹ However, they employed "spirals" drawings only.

In this paper, we extended the work of Pereira et al. [22] by presenting the following contributions: (i) a deeper analysis and explanation about the feature extraction process, as well as a tremor-based feature is also analyzed; (ii) we considered both spirals and meanders for the classification process; and (iii) we

¹ <http://www.fc.unesp.br/~papa/pub/datasets/Handpd/>.

also extended “HandPD” dataset with images and features from meanders. Since we are committed with science, we also made available to the readers this new dataset, and we believe it can serve as a basis for future research regarding Parkinson’s Disease diagnosis. The proposed approach is innovative in the sense that we can extract both the template and the drawings of each patient automatically, thus having no user intervention.

The remainder of this paper is organized as follows. Section 2 presents the methodology employed to design the dataset and the approach proposed to extract visual features from the handwriting exams. Section 3 states the experimental results and discussion, and Section 4 states conclusions and future works.

2. Materials and methods

In this section, we present the dataset built in this work, as well as the proposed approach to extract visual features from the exams.

2.1. HandPD dataset

The HandPD dataset was collected at the Faculty of Medicine of Botucatu, São Paulo State University, Brazil. It is composed of images extracted from handwriting exams of 92 individuals, divided in two groups: (i) the first one contains 18 exams of healthy people, named *control group*, with 6 male subjects and 12 female individuals; (ii) the second group contains 74 exams of people affected with Parkinson’s disease, named *patient group*, having 59 male and 15 female subjects. Therefore, 80.44% of the dataset is composed of patients, and 19.56% is composed of control individuals. Although the dataset is unbalanced, it is easier to achieve similar proportions by adding more control individuals than patients.

The control group is composed of 16 right-handed and 2 left-handed individuals, with an average age of 44.22 ± 16.53 years. In regard to the patient group, we have 69 right-handed and 5 left-handed individuals with an average age of 58.75 ± 7.51 years. Therefore, one can observe that the dataset is not age-biased, which provides an interesting scenario for learning purposes. In fact, most patients are considerably older than 60 years, since Parkinson’s disease usually gets worse within this age group. On the other hand, the dataset is heterogeneous enough to contain a 38-year-old male patient as well.

In order to compose the dataset, each subject is asked to fill a form in order to fulfill some task, such as drawing circles, spirals and meanders. Fig. 1a displays an exam of a 56-year-old male patient, in which we can observe the tremor inherent to Parkinson’s disease. Note that the patient is required to perform 6 distinct activities (Fig. 1a–f), which consist in the repetition of several operations in accordance with certain drawings. However, the analysis of the images will be focused on tasks c and d only, which are related to drawing 4 spirals and 4 meanders according to the template. Fig. 1b depicts an empty form, in which one can observe the templates regarding spirals and meanders.

After being filled out, the forms are digitized for the further extraction of spirals and meanders. Such step is performed by

hand, where each drawing is cropped to its minimum bounding box (or close to it). Soon after, the cropped spiral and meander images are numbered as follows: 1, 2, 3, 4 concerning the spirals from left to right, and 5, 6, 7, 8 concerning the meanders from left to right. Therefore, the entire dataset is composed of 736 images labeled in two groups: patients (296) and control (72). Also, the dataset comprises 368 images from each drawing, i.e., spirals and meanders. The reader can refer to the HandPD home-page for more technical details about organization of the dataset.

2.2. Feature extraction from visual description

In this section, we describe the methodology used to extract the features and keypoints from spiral and meander forms. In order to fulfill this task, we split the proposed methodology in two stages: (i) image preprocessing and (ii) the feature extraction. In the first stage (Section 2.2.1), we design an approach to automatically separate the handwritten trace (HT) from the exam template (ET), considering both spirals and meanders, since the images are not registered to each other. Soon after, in the second stage (Section 2.2.2), we used the HT and ET extracted from images to compute the visual features.

2.2.1. Handwritten trace and exam template

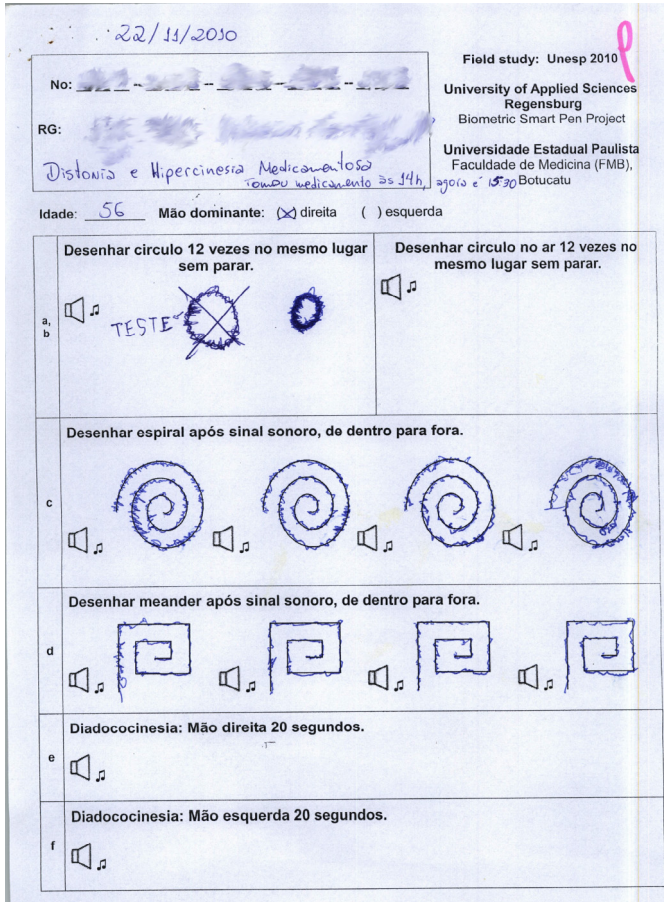
In order to extract both HT and ET, we merged some classical image processing techniques such as blurring filters and mathematical morphology, with the process of extracting either HT or ET contours performed separately. Since the images were digitized, we applied a preprocessing step to reduce noise and undesirable artifacts by means of a 5×5 mean filter.² Later on, we extracted the exam template by applying a thresholding in the smoothed image, aiming to obtain a binary mask $M_{ET}^i(I)$. This step is performed as follows:

$$M_{ET}^i(I) = \begin{cases} 0 & \text{if } R^i(I) < 100 \wedge G^i(I) < 100 \wedge B^i(I) < 100 \\ 1 & \text{otherwise,} \end{cases} \quad (1)$$

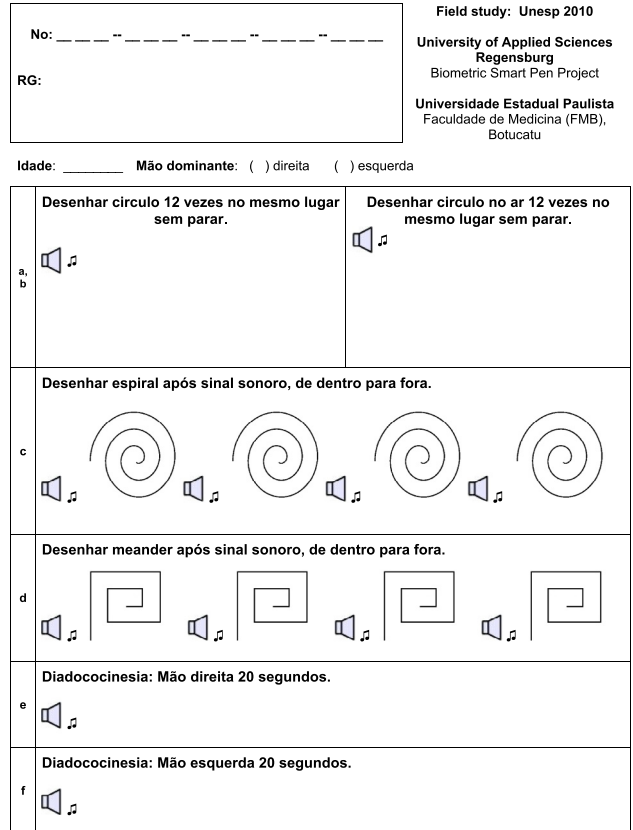
where $R^i(I)$, $G^i(I)$ and $B^i(I)$ stand for the value of pixel i of the input image I considering the channels “Red”, “Green” and “Blue”, respectively. If Equation 1 is satisfied, the foreground (ET) pixels will be set to 0 (“black” color), and the background pixels will be set to 1 (“white” color), as displayed in Fig. 3. Since the ET in the original image is supposed to be black or near-black (the original—empty—form is colorless), it is reasonable to assume low brightness values for such pixels when looking for the form itself. Finally, we applied an opening operation (erosion followed by a dilation) to guarantee a fully connected ET. Fig. 2 shows the proposed pipeline for the ET extraction.

In regard to the HT extraction step, we employed a similar methodology to the one used to extract the ET, but now with some additional steps and a different thresholding method, since both HT and the background are blue-colored due to the digitation process. First, we applied a 5×5 mean filter followed by a 5×5 median filter to smooth the image in order

² Notice that the size of this convolutional kernel was set up empirically.



(a)



(b)

Fig. 1 – Handwriting exams (a) filled out by a 56-year-old PD patient, and (b) an empty exam with the templates.

to reduce noise and small artifacts, mainly those around the HT's borders (once again, both filter sizes were determined empirically). Further, the filtered image F is thresholded using the following equation:

$$M_{HT}^i(F) = \begin{cases} 255 & \text{if } |R^i(F) - G^i(F)| < 40 \wedge |R^i(F) - B^i(F)| < 40 \wedge |G^i - B^i| < 40 \\ F^i & \text{otherwise,} \end{cases} \quad (2)$$

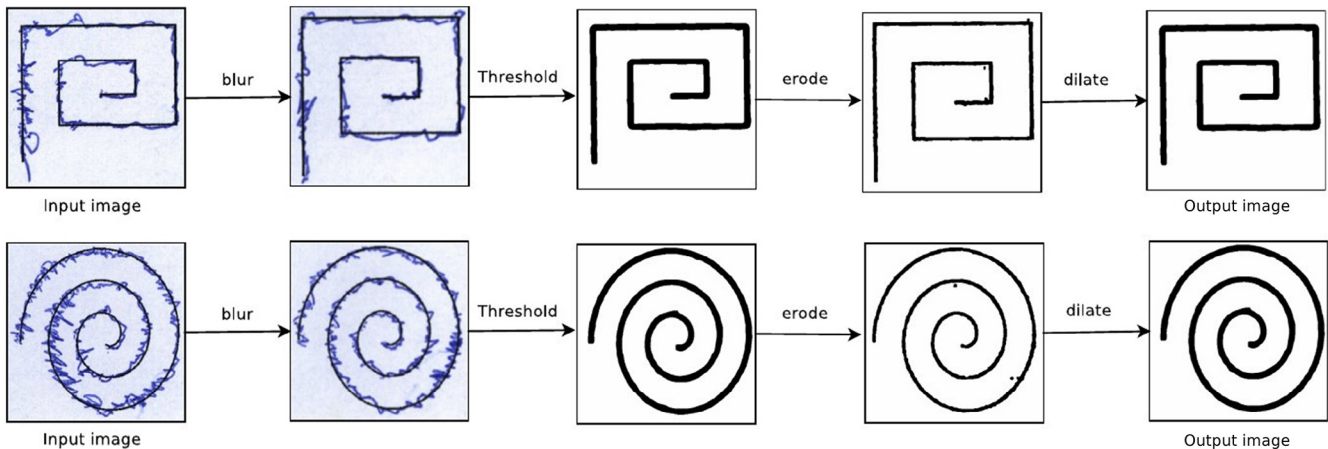


Fig. 2 – Image processing steps concerning ET extraction.

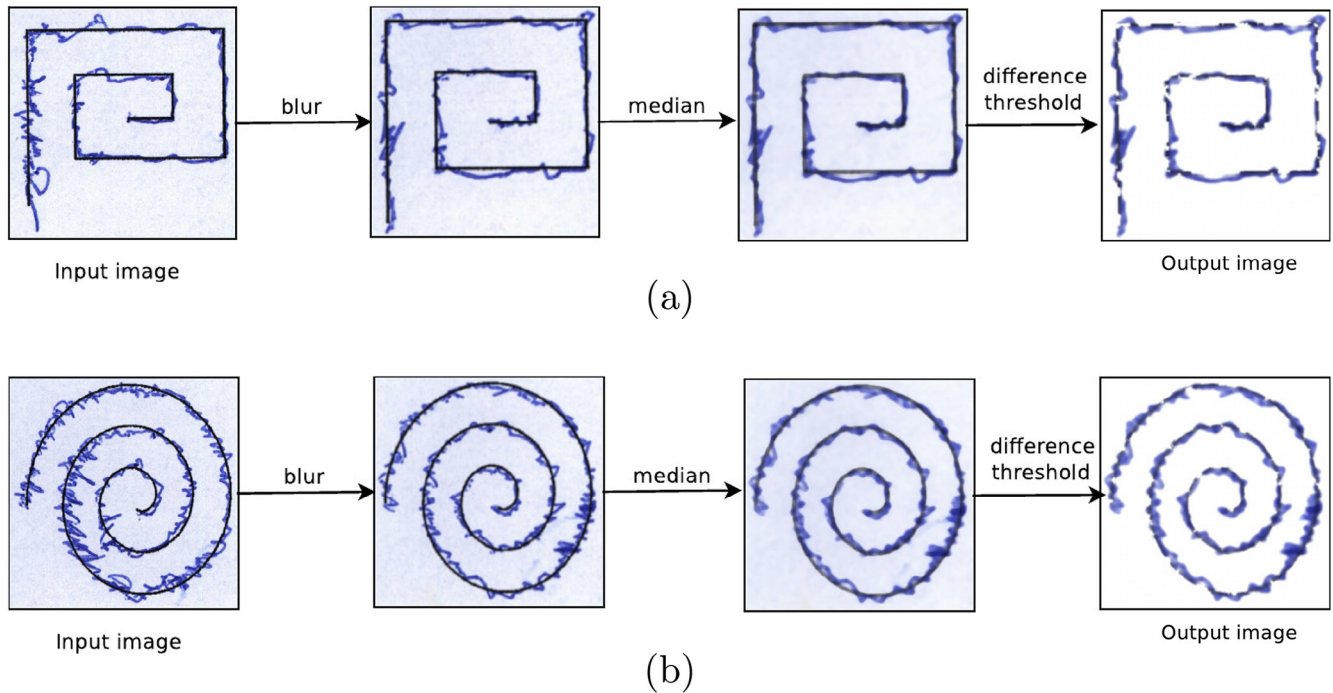


Fig. 3 – Image processing steps concerning HT extraction.

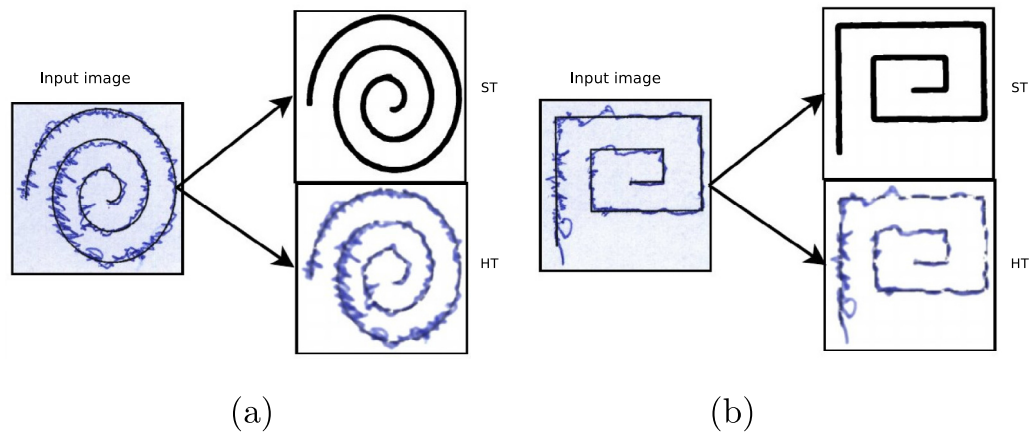


Fig. 4 – Spiral and meander images and their corresponding HT and ET extracted using the proposed methodology for a (a) spiral and a (b) meander.

where F^i represents the brightness of pixel i . The intuitive idea behind this step is to remove pixels with quasi-similar values for the three channels (i.e., background pixels), and to maintain pixels with considerable differences between the channels (foreground—HT—pixels). Fig. 3 shows the proposed pipeline for the HT extraction³. As a matter of fact, the fixed thresholds employed in this work obtained better results than some automatic approaches. Although they may not work well with

images acquired from a different procedure to the one used here, they seemed to be very suitable concerning the protocol adopted in this work.

Fig. 4 illustrates a spiral and a meander image and their corresponding ET and HT extracted using the proposed methodology. One can observe the quality of both template and trace extracted from the images.

2.2.2. Feature extraction

The feature extraction step aims at describing both HT and ET, and then to compare them in order to evaluate the “amount of difference” between the two images. In fact, this difference among images is computed over points sampled at the

³ Notice that the value 255 in Equation 2 stands for the triplet (255, 255, 255), since we have an RGB image as the result of thresholding operation.

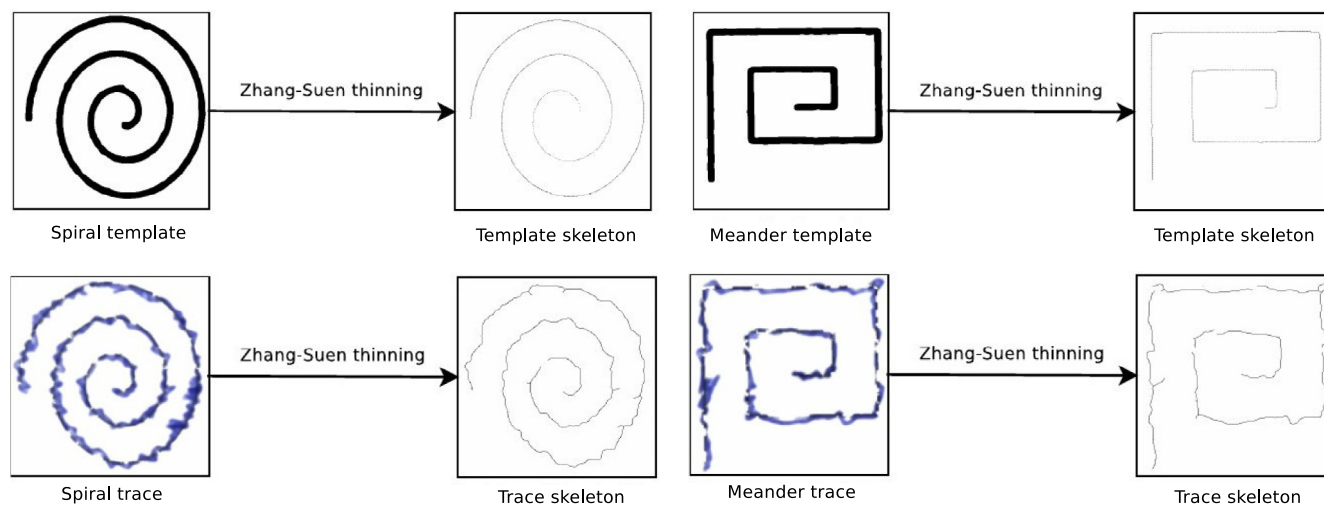


Fig. 5 – Thinning of HT and ET using Zhang–Suen algorithm.

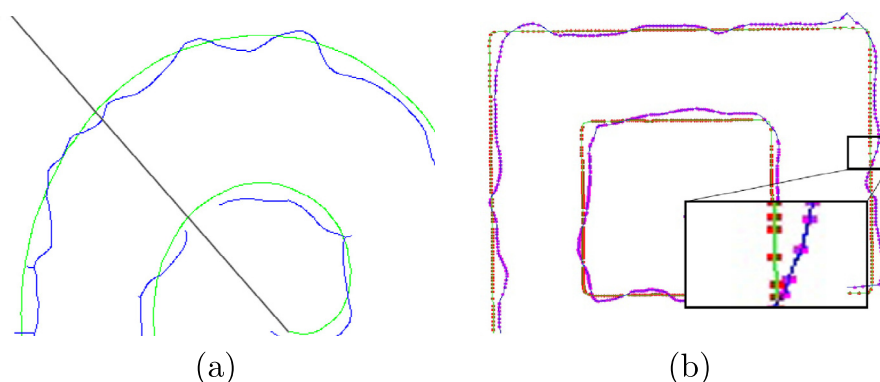


Fig. 6 – Sampling process: (a) a certain region with discontinuities, and (b) the proposed fair sampling process. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

very same positions considering HT and ET images. At each point, we extracted a set of features that will represent the whole template or handwritten trace. First, we need a concise and compact representation of both HT and ET, which is accomplished here by means of the skeleton of the thresholded images. Therefore, we extracted the skeleton of HT and ET images based on the Zhang–Suen thinning algorithm [23], which consists of two parallel routines: (i) to remove the south-east boundary points and the north-west corner points, and (ii) to remove the north-west boundary points and the south-east corner points. Fig. 5 depicts the thinning result of the spiral and meander templates, as well as the handwritten trace.

Even after the pre-processing step, the template and handwritten images may contain small discontinuities (blue lines in Fig. 6(a)). Therefore, we need to select the sample points from the template and handwritten spiral/meander very carefully. As such, points in regions that contain discontinuities should be discarded. This phase is crucial, since it has a considerable influence in the feature extraction step, which may affect the learning process as well.

In regard to the selection of sampled points, we trace 360 rays⁴ from the center of the spiral/meander to the image borders. For this task, we created two empty lists: (i) template points and (ii) handwritten points. The ray tracing process begins from the more external point of the spiral or the meander. For each ray, we capture its intersections with the template and the handwritten trace, and if this ray intercepts only one of the images, this point is discarded; otherwise, the pair of points is inserted in their respective list of points (template or handwritten). Therefore, with the aforementioned procedure, we can guarantee a fair sampling by considering only points presented in both images. Fig. 6(b) shows a thinned meander with overlapped traces (template and handwritten), as well as the highlighted sampling points obtained by means of the proposed fair sampling process.

⁴ Notice that the value 360 was obtained empirically, since this amount of sampling points has showed a good trade-off between efficiency and accuracy.

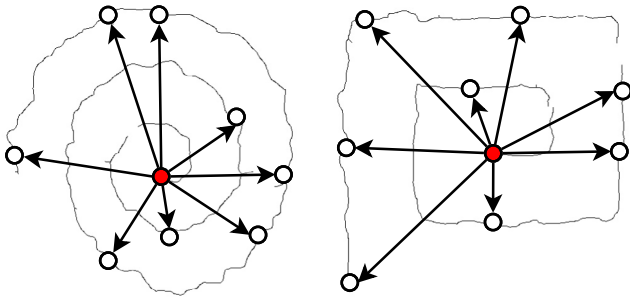


Fig. 7 – Some random points and the straight lines representing their connections with the spiral’s and meander’s center point. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Further in the sampling process, we then extract nine numeric features from each skeleton (i.e., HT and ET) by measuring the statistical differences between them. However, prior to the feature description, we introduce to the reader the definition of “radius” of a spiral or meander point, which is basically the length of the straight line that connects this point to the center of the spiral or meander, as displayed in Fig. 7. The “red” point stands for the spiral’s or meander’s center, being some random (“white”) points connected to the thinned spiral (skeleton) through the arrows with straight lines.

A brief description of each feature is given below:

- f_1 : Root Mean Square (RMS) of the difference between HT and ET radius. The RMS is computed as follows:

$$\text{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^n (r_{\text{HT}}^i - r_{\text{ET}}^i)^2}, \quad (3)$$

where n is the number of sample points drawn from each HT and ET skeleton, and r_{HT}^i and r_{ET}^i denote the HT and ET radius considering the i -th sampled point, respectively.

- f_2 : the maximum difference between HT and ET radius, i.e.:

$$\Delta_{\max} = \operatorname{argmax}_i \{|r_{\text{HT}}^i - r_{\text{ET}}^i|\}; \quad (4)$$

- f_3 : the minimum difference between HT and ET radius, i.e.:

$$\Delta_{\min} = \operatorname{argmin}_i \{|r_{\text{HT}}^i - r_{\text{ET}}^i|\}; \quad (5)$$

- f_4 : the standard deviation of the differences between HT and ET radius;
- f_5 : Mean Relative Tremor (MRT): Pereira et al. [22] proposed this quantitative evaluation to measure the “amount of tremor” of a given individual’s HT, being defined as the mean difference between the radius of a given sample and its d left-nearest neighbors. The MRT is computed as follows:

$$\text{MRT} = \frac{1}{n-d} \sum_{i=d}^n |r_{\text{ET}}^i - r_{\text{ET}}^{i-d+1}|, \quad (6)$$

where d is the displacement of the sample points used to compute the radius difference.⁵ The following three features are computed based on the relative tremor $|r_{\text{ET}}^i - r_{\text{ET}}^{i-d+1}|$;

- f_6 : the maximum ET;
- f_7 : the minimum ET;
- f_8 : the standard deviation of ET values;
- f_9 : the number of times the difference between HT and ET radius changes from negative to positive, or vice-versa.

Finally, the features were normalized as follows:

$$f'_i = \frac{f_i - \mu_i}{\sigma_i}, \quad (7)$$

where f'_i denotes the normalized version of feature f_i , and μ_i and σ_i stand for the average and standard deviation of feature f_i , $i = 1, 2, \dots, 9$.

3. Experiments and results

In this section, we present the experimental results to access the robustness of the proposed dataset and feature extraction approach.⁶ Also, we evaluate three pattern recognition techniques: Naïve Bayes (NB), Optimum-Path Forest (OPF), and Support Vector Machines with Radial Basis Function (SVM-RBF). Note that the kernel parameters concerning SVM are optimized through cross-validation. In regard to OPF, we used LibOPF [24], and with respect to NB and SVM, we used scikit-learn [25].

In order to evaluate the proposed approach, we performed three different rounds of experiments. The first one (Section 3.1) uses 75% of the dataset for training purposes and the remaining 25% for the classification phase. However, instead of partitioning the dataset randomly, we created four subsets in order to guarantee that each individual will be represented in the dataset with its 3 spirals/meanders, with the remaining one being used for classification purposes. In this experiment, the spiral- and meander-based datasets are used individually. In the second experiment (Section 3.2), we decided to conduct a cross-validation procedure with 20 runnings. Now, we no longer guarantee each individual will be represented in both training and test sets. In the third round (Section 3.3), we conducted some experiments in order to check whether we can benefit from the learning process over spirals and meanders in one single approach, i.e. by using them together. Finally, we present a discussion about the experiments, as well as some insights about this research.

3.1. Experiment 1

Since each individual contains four spirals/meanders in the datasets, we employed a constrained hold-out approach to

⁵ In this work, we used $d = \{1, 3, 5, 7, 10, 15, 20\}$, with $d = 10$ being the one that maximized the PD recognition rate.

⁶ The proposed dataset and the extracted features are available at <http://www.fc.unesp.br/~papa/pub/datasets/Handpd>.

Table 1 – Experimental results considering the spiral-based dataset.

	OPF	NB	SVM
Control group	31.94% ± 5.32	62.50% ± 5.32	2.78% ± 5.56
Patient group	76.35% ± 3.22	69.26% ± 7.18	99.66% ± 0.68
Global	54.15% ± 3.58	65.88% ± 4.57	51.22% ± 2.91

guarantee that each of them will be represented in both training and testing sets concerning both spiral- and meander-based datasets. Tables 1 and 2 display the mean recognition rates considering all four configurations of training and test sets for the spiral- and meander-based datasets, respectively. One can observe that NB obtained the best global results concerning Spiral dataset, while SVM achieved the best results over Meander dataset. Notice that the global accuracy is the one proposed by Papa et al. [15], which considers unbalanced datasets, while the recognition rates per class (i.e. Control and Patient groups) are computed using the standard approach (the ratio between correct classifications and the total number of samples for that specific class). The values in bold stand for the most accurate ones considering the standard deviation only. As aforementioned, in this round of experiments we used a similar approach to a 4-fold cross-validation, but we ensured that we have three drawings for the same individual for training purposes. Therefore, we can guarantee all individuals are represented in both training and test sets. However, as we have only four accuracy values to compute the mean recognition rates and their standard deviation, we did not employ any robust statistical evaluation in this experiment.

Curiously, a different behavior considering each classifier and both datasets can be observed. Note that OPF obtained better results over meander dataset when compared to the spirals dataset, while NB holds the opposite situation. Such situation motivated us to consider a bag-of-classifiers in order to check whether a combination among all classifiers will make the results better or not (Section 3.3).

3.2. Experiment 2

In this section, we consider a cross-validation procedure with 20 runnings to assess the robustness of the proposed approach under a different scenario. Therefore, we can no longer guarantee each individual will be represented in both training and test sets, but we can obtain more conclusive results by means of the Wilcoxon signed-rank statistical test [26]. In this work, we used a significance of 0.05. Tables 3 and 4 present the mean recognition rates considering spiral- and meander-based datasets, respectively. Once again, we can observe results very similar to the ones obtained in the previous section. The

Table 2 – Experimental results considering the meander-based dataset.

	OPF	NB	SVM
Control group	34.72% ± 8.33	20.83% ± 41.67	36.11% ± 9.62
Patient group	85.81% ± 4.20	79.73% ± 33.34	96.62% ± 2.59
Global	60.27% ± 4.02	50.28% ± 4.18	66.37% ± 4.01

Table 3 – Average results considering the spiral-based dataset and a cross-validation with 20 runnings.

	OPF	NB	SVM
Control group	26.39% ± 9.17	65.56% ± 11.48	1.67% ± 4.07
Patient group	78.58% ± 5.02	62.91% ± 12.65	98.65% ± 4.34
Global	52.48% ± 5.32	64.23% ± 7.11	50.16% ± 1.71

values in bold stand for the most accurate techniques considering the aforementioned statistical evaluation.

Since the dataset is dominated by patients, all classifiers achieved better recognition rates for that class, except for NB considering spiral and meander datasets. In fact, with respect to this classifier, the accuracy rates per class were similar to each other considering the spiral dataset, but considerably distinct with respect to meanders. NB seemed to better manage control individuals, but we can also observe the highest standard deviations for this classifier as well.

3.3. Experiment 3

In this section, we conducted an experiment to check whether we can benefit from information learned from both drawings. We used a standard majority voting for each classifier, and in case of ties, we opted to use the classification given by the meanders dataset, since it has been the most accurate (Section 3.2). Table 5 presents the mean accurate rates for each class, as well as the global accuracy. Notice that we used the very same sets employed in the first experiment (Section 3.1), since we can guarantee that both spiral and meander analyzed at a given time step of the classification algorithm come from the same individual.

The results evidenced that one may not benefit from the combined information between spirals and meanders, since the results are now worse than the ones obtained with meanders only. The main problem is related to the inconsistency among samples from the control and patient groups. That means we can not observe that different drawings can help each other since we have inconsistencies at the very same exam for different patients. In the next section we discuss such statements in more details.

Table 4 – Average results considering the meander-based dataset and a cross-validation with 20 runnings.

	OPF	NB	SVM
Control group	32.78% ± 12.08	80.83% ± 16.37	36.94% ± 10.71
Patient group	82.30% ± 3.72	37.57% ± 22.83	96.49% ± 2.50
Global	57.54% ± 6.35	59.20% ± 4.78	66.72% ± 5.33

Table 5 – Average results considering the combination process between spirals and meanders using the constrained 4-fold approach.

	OPF	NB	SVM
Control group	64.96% ± 16.29	27.30% ± 37.36	12.50% ± 25.00
Patient group	60.23% ± 4.73	70.36% ± 39.08	96.49% ± 2.50
Global	55.86% ± 3.63	45.79% ± 4.15	58.61% ± 2.84

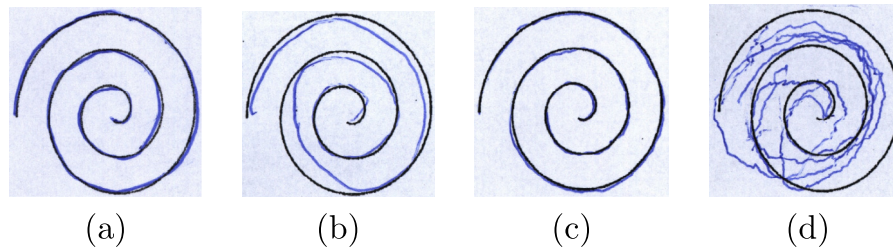


Fig. 8 – Spirals from the control group (a, b) and from the patient group (c, d).

3.4. Discussion

The experiments conducted in this paper may drive us to three main conclusions: (i) first, to ensure that we have the very same patient in both training and testing sets does not seem to benefit the final classification rates, since the results obtained in Sections 3.1 and 3.2 were very similar to each other; (ii) second, meanders can provide more reliable recognition rates; and (iii) finally, it seems the combination of information provided by both spirals and meanders does not benefit the final classification rates.

The main problem related to PD automatic recognition is the patients in the initial stage of the disease, since they often do not present any symptoms related to tremors. Fig. 8 depicts some examples of spirals from both control and patient groups. If we consider Fig. 8b and 8c, for instance, the former belongs to a control individual, and the latter belongs to a patient. Clearly, the patient exam looks like from someone who is not affected by the disease, i.e., it is very similar to Fig. 8a. The high variability of the dataset may lead the classifiers to errors as well. However, the main idea in designing such dataset is to capture such sort of problems, which are not straightforward to solve. Obviously, Fig. 8d is easier to be labeled as patient than Fig. 8c, but the opposite situation is not true. Actually, the main problem is when Fig. 8c is represented in the training set, not in the test set. In the former situation, this exam has a high probability to be an outlier, thus leading the learning process to mistakes in the classification phase. The latter situation usually only affects that sample only, i.e. it will be probably labeled as control. Although it may decrease the overall classification rate, the major problem is related to the fact that such exam will be a false negative, thus postponing the treatment of the disease.

Fig. 9 displays some meanders from both control and patient groups. A similar situation to the one faced with spirals can

also be observed with meanders. The high variability of the dataset makes the classifiers more prone to errors, thus turning the problem of identifying PD in the early stages quite complicated. However, the proposed approach obtained $\approx 67\%$ of recognition rates using meanders, which we consider a very suitable result. As aforementioned, we have not noticed any particular image-based dataset available in the internet, as well as with the proposed pipeline for feature extraction adopted in this work.

4. Conclusion

In this paper, we dealt with the problem of Parkinson's Disease recognition by combining machine learning and computer vision techniques. The main contributions are related to the design of a new dataset that contains images from both spirals and meanders, which are cropped out from digitized handwritten exams, and we proposed a pipeline that can deal with the problem of learning from non-registered images. The proposed approach can automatically extract both the template and the handwritten trace from each exam for further feature extraction and classification.

The experimental results can lead us to conclude that meanders are more informative than spirals, since the latter pose a greater challenge due to the contours inherent to their shape. Also, the combination of both approaches did not seem to improve the results. The main problem is related to the high variability of the dataset, which comprises patients at the very early stages of the disease, thus being very difficult to be diagnosed. In regard to future works, we intend to increase the dataset with more samples from the control group, as well as to design new features that can better distinguish between control individuals and patients.

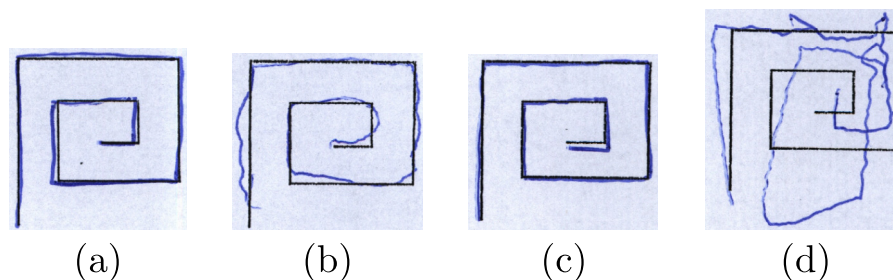


Fig. 9 – Meanders from the control group (a, b) and from the patient group (c, d).

Acknowledgment

The authors are grateful to CAPES PROCAD 2966/2014 grant, FAPESP grants #2009/16206-1, #2013/20387-7 and #2014/2014/16250-9, as well as CNPq grants #303182/2011-3, #70571/2013-6 and #306166/2014-3.

REFERENCES

- [1] R.E. Burke, Evaluation of the Braak staging scheme for Parkinson's disease: introduction to a panel presentation, *Mov. Disord.* 25 (S1) (2010) S76–S77.
- [2] J. Parkinson, An essay on the shaking palsy, *J. Neuropsychiatry Clin. Neurosci.* 14 (2) (1817) 223–236.
- [3] B.E. Sakar, M.E. Isenkul, C.O. Sakar, A. Sertbas, F. Gurgun, S. Delil, et al., Collection and analysis of a parkinson speech dataset with multiple types of sound recordings, *IEEE J. Biomed. Health Inform.* 17 (2013) 828–834.
- [4] M.A. Little, P.E. McSharry, E.J. Hunter, J. Spielman, L.O. Ramig, Suitability of dysphonia measurements for telemonitoring of parkinson's disease, *IEEE Trans. Biomed. Eng.* 56 (4) (2009) 1015–1022.
- [5] J.C. Pereira, A.O. Schelp, A.N. Montagnoli, A.R. Gatto, A.A. Spadotto, L.R. Carvalho, Residual signal auto-correlation to evaluate speech in Parkinson's disease patients, *Arq. Neuropsiquiatr.* 64 (4) (2006) 912–915.
- [6] A.K. Ho, R. Lansek, C. Maricliani, J.L. Bradshaw, S. Gates, Speech impairment in a large sample of patients with parkinson's disease, *Behav. Neurol.* 3 (11) (1998) 131–137.
- [7] J.A. Logemann, H.B. Fisher, B. Boshes, E.R. Blonsky, Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of parkinson patients, *J. Speech Hear. Disord.* 43 (11) (1978) 47–57.
- [8] S. Zhao, F. Rudzicz, L.G. Carvalho, C. Marquez-Chin, S. Livingstone, Automatic detection of expressed emotion in parkinson's disease, in: *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4813–4817, 2014.
- [9] A. Tsanas, M.A. Little, P.E. McSharry, J. Spielman, L.O. Ramig, Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease, *IEEE Trans. Biomed. Eng.* 59 (5) (2012) 1264–1271.
- [10] B. Harel, M. Cannizzaro, P.J. Snyder, Variability in fundamental frequency during speech in prodromal and incipient parkinson's disease: a longitudinal case study, *Brain Cogn.* 6 (1) (2004) 24–29.
- [11] T.E. Eichhorn, T. Gasser, N. Mai, C. Marquardt, G. Arnold, J. Schwarz, et al., Computational analysis of open loop handwriting movements in parkinson's disease: a rapid method to detect dopamimetic effects, *Mov. Disord.* 11 (3) (1996) 289–297.
- [12] S. Rosenblum, M. Samuel, S. Zlotnik, I. Erikk, I. Schlesinger, Handwriting as an objective tool for parkinson's disease diagnosis, *J. Neurol.* 260 (9) (2013) 2357–2361.
- [13] 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.
- [14] A.A. Spadotto, R.C. Guido, J.P. Papa, A.X. Falcão, Parkinson's disease identification through optimum-path forest, in: *International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6087–6090, 2010.
- [15] J.P. Papa, A.X. Falcão, C.T.N. Suzuki, Supervised pattern classification based on optimum-path forest, *Int. J. Imag. Syst. Technol.* 19 (2) (2009) 120–131.
- [16] J.P. Papa, A.X. Falcão, V.H.C. Albuquerque, J.M.R.S. Tavares, Efficient supervised optimum-path forest classification for large datasets, *Pattern Recognit.* 45 (1) (2012) 512–520.
- [17] A.A. Spadotto, R.C. Guido, F.L. Carnevali, A.F. Pagnin, A.X. Falcão, J.P. Papa, Improving parkinson's disease identification through evolutionary-based feature selection, in: *International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 7857–7860, 2011.
- [18] F.S. Gharehchopogh, P. Mohammadi, Article: a case study of parkinsons disease diagnosis using artificial neural networks, *Int. J. Comput. Appl.* 73 (19) (2013) 1–6.
- [19] S. Pan, S. Iplikci, K. Warwick, T.Z. Aziz, Parkinson's disease tremor classification, a comparison between support vector machines and neural networks, *Expert Syst. Appl.* 19 (2012) 10764–10771.
- [20] M. Hariharan, K. Polat, R. Sindhu, A new hybrid intelligent system for accurate detection of parkinson's disease, *Comput. Methods Programs Biomed.* 11 (3) (2014) 904–913.
- [21] M. Peker, B. Sen, D. Delen, Computer-aided diagnosis of parkinson's disease using complex-valued neural networks and mRMR feature selection algorithm, *J. Healthc. Eng.* 6 (3) (2015) 281–302.
- [22] C.R. Pereira, D.R. Pereira, F.A. da Silva, C. Hook, S.A.T. Weber, L.A.M. Pereira, et al., A step towards the automated diagnosis of parkinson's disease: analyzing handwriting movements, in: *IEEE 28th International Symposium on Computer-Based Medical Systems*, pp. 171–176, 2015.
- [23] T.Y. Zhang, C.Y. Suen, *A Fast Parallel Algorithm for Thinning Digital Patterns*, ACM, New York, NY, USA, 1984.
- [24] J. Papa, C. Suzuki, A. Falcao, LibOPF: A library for the design of optimum-path forest classifiers, software version 2.1. <<http://www.ic.unicamp.br/afalcao/libopf/index.html>>, 2014 (accessed 01.02.2016).
- [25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., Scikit-learn: machine learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- [26] F. Wilcoxon, Individual comparisons by ranking methods, *Biomet. Bull.* 1 (6) (1945) 80–83.