A Step Towards the Automated Diagnosis of Parkinson's Disease: Analyzing Handwriting Movements

Clayton R. Pereira
Federal University - UFSCAR
São Carlos, Brazil
Email: clayton.pereira@dc.ufscar.br

Danillo R. Pereira, Francisco A. da Silva
Western University - UNOEST
Presidente Prudente, Brazil
Email: dpereira@ic.unicamp.br

Christian Hook
Ostbayerische Technische Hochschule
Regensburg - Germany
Email: christian.hook@hs-regensburg.de

Silke A. T. Weber, Luis A. M. Pereira, João P. Papa
São Paulo State University - UNESP
São Paulo, Brazil
Email:silke@fmb.unesp.br,
luismartinspr@gmail.com, papa@fc.unesp.br

***Abstract*—Parkinson's disease (PD) has affected millions of people world-wide, being its major problem the loss of movements and, consequently, the ability of working and locomotion. Although we can find several works that attempt at dealing with this problem out there, most of them make use of datasets composed by a few subjects only. In this work, we present some results toward the automated diagnosis of PD by means of computer vision-based techniques in a dataset composed by dozens of patients, which is one of the main contributions of this work. The dataset is part of a joint research project that aims at extracting both visual and signal-based information from healthy and PD patients in order to go forward the early diagnosis of PD patients. The dataset is composed by handwriting clinical exams that are analyzed by means of image processing and machine learning techniques, being the preliminary results encouraging and promising. Additionally, a new quantitative feature to measure the amount of tremor of an individual's handwritten trace called Mean Relative Tremor is also presented.**

***Keywords*-Parkinson's disease, machine learning, movement disorders**

I. Introduction

Parkinson's disease (PD) is a neurological illness that affects a person's movements, and may cause tremors, slowness of movement, muscle stiffness and imbalance, as well as changes in speech and writing skills [1]. PD was first described by the English physician James Parkinson [2], being classified as a degenerative, chronic and progressive disease. Although several PD symptoms are well-known, it is still unheard-of a trivial test to diagnose it, at least accurately enough. Additionally, it is not straightforward to establish the PD level soon after its diagnosis.

In the past few years, several works have attempted at solutions to aid the PD diagnosis. Among them, expert systems based on machine learning techniques have shown promising results [3]. Most part of works performs signal analysis of the patient voice [4], [5], since its voice capability is also gradually compromised by PD. In Little et al. [4], a dataset was designed through a cooperation between the University of Oxford and the National Centre for Voice and Speech (Denver, Colorado). The dataset was composed of biomedical voice measurements from 31 male and female subjects, being 23 patients diagnosed with PD, and with ages ranging from 46 to 85 years (average age of 65.8, and standard deviation of 9.8) [14].

Micrography is another widely approach used for the diagnosis of a patient with Parkinson's disease [6], which is basically a writing exam. Such technique is considered an objective measure for characterizing the disease, since a PD patient possibly features the reduction of the size of calligraphy and hand tremor, which are currently identified as being the main symptoms of Parkinson's disease. Nowadays, this procedure is often conducted by filling out forms.

Pattern recognition techniques have also been applied to automatic PD recognition. Spadotto et al. [7], for instance, introduced the Optimum-Path Forest [8], [9] (OPF) classifier to the aforementioned context. Gharehchopogh et al. [10] used Artificial Neural Networks with Multi-Layer Perceptron (ANN-MLP) to diagnose the effects caused by Parkinson's disease. Pan et al. [11] analyzed the performance of Support Vector Machines (SVM), ANN-MLP and Neural networks with Radial Basis Function (RBF) to compare the onset of tremor in patients with Parkinson's disease.

Zuo et al. [12] proposed a computer-oriented system to aid PD diagnosis based on Particle Swarm Optimization and fuzzy k-nearest neighbors classifier. Such techniques have been used for parameter optimization and feature selection simultaneously, achieving an accuracy of 97.47%. Chao et al. [13] employed a hybrid method

2372-9198/15 $3.00 © 2015 IEEE
DOI 10.1109/CBMS.2015.34

based on Subtractive Clustering Features Weighting and Extreme Learning Machines for PD identification as well. The authors stated a classification accuracy and sensitivity of 99.49% and 100%, respectively. It is worth mentioning that both techniques made use of the aforementioned dataset described in [14]. Although the results reported are outstanding, 74% of the dataset comprises patients with Parkinson's disease, which may bias the results.

However, most part of works employed signal-based features only. The reader can face just a few of them that made use of images to aid the task of PD recognition. Haller et al. [15], for instance, presented a system to help the diagnosis of Parkinson's disease using Magnetic Resonance Images (MRI), since this type of image can provide important information about the process of connection between different brain regions. The main assumption is related that PD patients may have brain regions that are strongly affected, and therefore can be visually identified.

Nonetheless, MRI images may be too costly to be obtained, and also require the patient to stand still for the image acquisition process, which may not be straightforward for PD patients. Thereafter, in this paper we propose a much cheaper approach to aid the diagnosis of Parkinson's disease by means of image-based features. We designed a dataset composed by handwriting exams in some specific tasks to asses the writing skills of a given patient, being such exam further digitized for the application of computer vision techniques. After that, shape-based features are then extracted from those images and used to feed supervised pattern recognition techniques. Another key contribution of this work is to make available such dataset composed by dozens of patients, since most datasets used out there contain a few samples (patients) only, as well as we proposed a new feature called "Mean Relative Tremor", which aims at measuring the "amount of tremor" considering the handwriting trace of the individual. In short, the main contributions of this paper are three fold: (i) to design a dataset called "HandPD" composed by dozen of samples and to make it public, (ii) to propose an image processing pipeline for feature extraction, and (iii) to present a new quantitative description about the amount of tremor of a given individual. The effectiveness of the proposed dataset is evaluated by means of some state-of-the-start supervised techniques, such as OPF, SVM and Naïve Bayes classifier (NB), being the results very promising.

The reminder of this paper is organized as follows. Section II presents the raw data used to design the dataset, and Section III describes the proposed approach to extract visual features of the handwriting exams. Section IV states the experimental results, and Section V presents conclusions and final remarks.

II. Parkison's Disease Assessment

One of the most challenging tasks when dealing with PD diagnosis is whether to use visual and/or signal-based information from patient exams. As aforementioned, previous works have used high-end image technology (MRI) for such purposes, but being expensive and may be invasive enough to the patient as well. Additionally, most signal-based datasets for PD recognition are small and biased, which may not reflect the real world.

In order to overcome such shortcomings, we developed a new dataset composed of images extracted from handwriting exams of 55 individuals, being 37 of them affected with Parkinson's disease (*patients group*), and the remaining 18 stand for healthy people (*control group*). Additionally, the ages of all individuals from the patient and control groups fall within the ranges $[38, 78]$ and $[19, 79]$, respectively. The patient group is divided into 25 male and 12 female individuals, whereas the control group contains 6 male and 12 female individuals. This dataset was collected at Faculty of Medicine of Botucatu, São Paulo State University, Brazil. The writing exam consists, essentially, in filling out a form in order to fulfil some tasks, such as drawing circles, spirals and meanders. Figure 1 displays an exam of a 56 years-old male patient, in which we can observe the tremor inherent to Parkinson's disease.

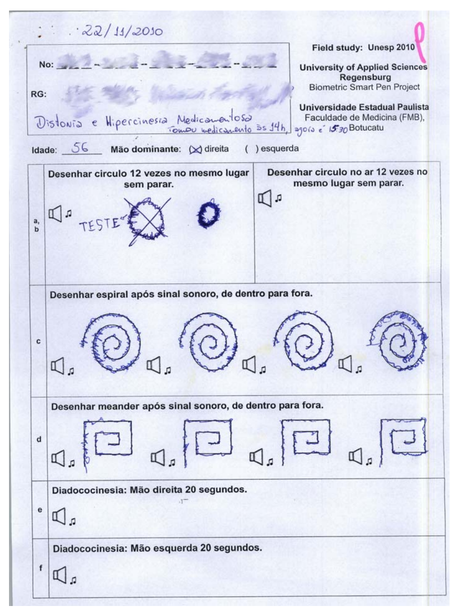

Figure 1. Handwriting exam filled out by a 56-years old PD patient.

Although we have tasks related to spirals and meanders, in this first version of the dataset we are considering spiral images only. Albeit we have four 4 spirals per exam, some individuals did not fill out all of them. Thus, the proposed dataset is composed of 373 samples (~ 4 spiral samples per individual). Figure 2 displays some examples of spirals extracted from the dataset.

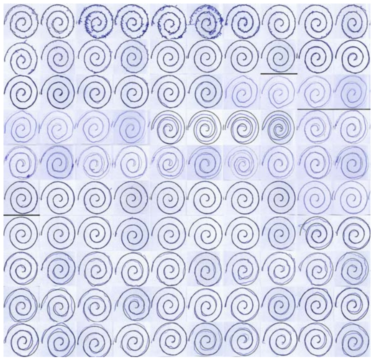

Figure 2. Some examples of spirals extracted from the dataset.

III. Extracting Knowledge from Handwritten Exams

In this section, we describe the methodology to design the proposed dataset from handwriting exams. In order to fulfil this task, we divide the proposed approach into two parts: (i) image processing, and (ii) feature extraction. First of all, we need to extract the handwritten trace (HT) from the spiral template (ST), since the images are not registered to each other. Soon after, we use both HT and ST to compute the features based on their shapes. Figure 3 depicts an example image and its corresponding handwritten trace and spiral template.

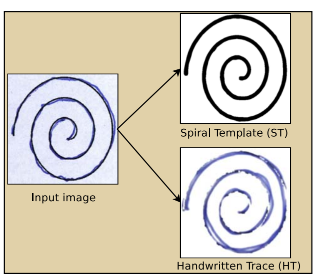

Figure 3. Example image and its corresponding handwritten trace and spiral template.

A. Spiral Template and Handwritten Trace Segmentation

In order to segment both ST and HT, we combined some classical image processing techniques such as blurring filters and mathematical morphology, being the process of extracting either ST and HT contours performed separately. Since the images were digitized, we applied a preprocessing step to reduce noise and undesirable artefacts by means of a 5×5 mean filter[1]. Later on, we extracted the spiral template by thresholding the smoothened image in order to obtain a binary mask M_{ST}^{i} of the spiral itself. This step is accomplished as follows:

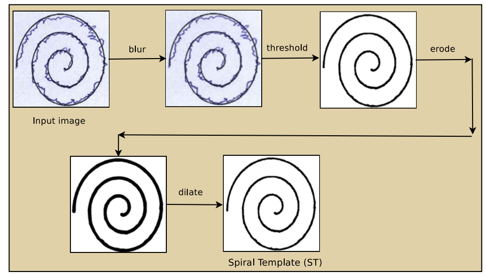

Figure 4. Image processing steps concerning the spiral template extraction.

$$M_{ST}^{i} = \begin{cases} 0 & \text{if } R^i < 100 \wedge G^i < 100 \wedge B^i < 100 \\ 1 & \text{otherwise,} \end{cases} \tag{1}$$

where R^i, G^i and B^i stand for the value of pixel i of the input image considering the channels "Red", "Green" and "Blue", respectively. In short, if Equation 1 is satisfied, the foreground (spiral template) pixels will be set to 1 ("black" color), and the background pixels will be set to "white" color, as displayed in Figure 3. Since the spiral template in the original (input) image is suppose to be black/near-black, it is reasonable to assume low brightness values for such pixels when looking for the spiral itself. Finally, we applied an opening operation (erosion followed by a dilation) to guarantee a fully connected spiral template. Figure 4 shows the proposed pipeline for the ST extraction.

In regard to the HT extraction step, we employed a similar methodology to the previous one used to extract the spiral template, but now with some additional steps and a different thresholding method, since both HT and the background are blue-colored. Firstly, we applied a 5×5 mean filter followed by a 5×5 median filter to smooth the image in order to reduce noise and small artefacts, 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:

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

[1]Notice the size of this convolutional kernel was set up empirically.

where F^i stands for 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 maintains pixels with considerable differences between the channels (foreground - HT - pixels). Figure 5 shows the proposed pipeline for the HT extraction[2].

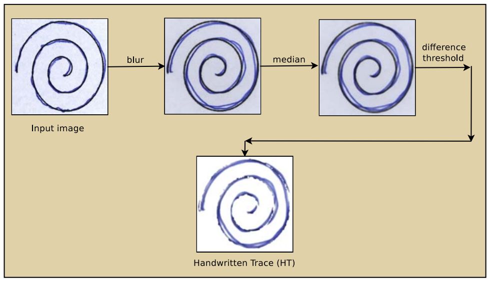

Figure 5. Image processing steps concerning the handwritten trace extraction.

B. Feature Extraction

The feature extraction step aims at describing both ST and HT, and then to compare them in order to evaluate the "amount of difference" between both images. In order to fulfil this task, we first need a concise and compact representation of both ST and HT, which is accomplished here by means of the skeleton of the thresholded images. Therefore, we extracted the skeleton of ST and HT images based on the Zhang-Suen thinning algorithm [17], 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. Figures 6 and 7 depict the thinning result of the spiral template and the handwritten trace, respectively.

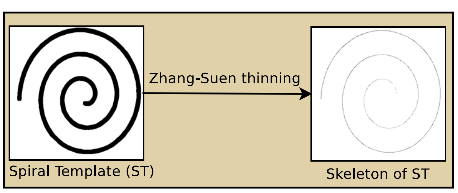

Figure 6. Thinning of spiral template using Zhang-Suen algorithm.

After that, we extracted nine numeric features from each skeleton (i.e., ST and HT) 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 point, which is basically

[2]Notice 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.

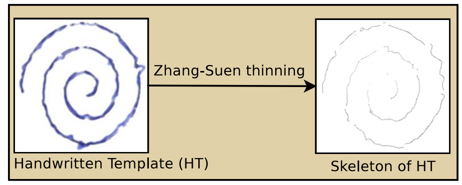

Figure 7. Thinning of handwritten trace using Zhang-Suen algorithm.

the length of the straight line that connects this point to the center of the spiral, as displayed in Figure 8. The "red" point stands for the spiral's center, being some random ("white") points connected to the thinned spiral (skeleton) through the red dashed straight lines.

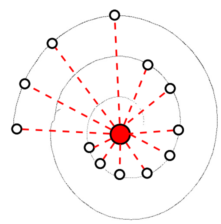

Figure 8. Some random points and the straight lines representing their connections with the spiral's center point.

A brief description of each feature is given below:

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

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

where n is the number of sample points drawn from each ST and HT skeleton, and r_{ST}^i and r_{HT}^i denote the ST and HT radius considering the i-th sampled point, respectively[3].

2) the maximum difference between ST and HT radius, i.e.:

$$\Delta_{max} = \underset{i}{\text{argmax}}\{|r_{ST}^i - r_{HT}^i|\}; \qquad (4)$$

3) the minimum difference between ST and HT radius, i.e.;

$$\Delta_{min} = \underset{i}{\text{argmin}}\{|r_{ST}^i - r_{HT}^i|\}; \qquad (5)$$

4) the standard deviation of the differences between ST and HT radius;

[3]In this paper, we set $n = 360$ due to empirical reasons, since this amount of sampling points has showed a good trade-off between efficiency and accuracy.

5) Mean Relative Tremor (MRT): we proposed this quantitative evaluation to measure the "amount of tremor" of a given individual's handwritten trace, being defined as the mean difference between the radius of a given sample and its d left-nearest neighbours. The MRT is computed as follows:

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

where n is the number of sample points, and d is the displacement of the sample points used to compute the radius difference[4]. The following three features are computed based on the relative tremor $|r_{HT}^{i} - r_{HT}^{i-d+1}|$;

6) the maximum HT;
7) the minimum HT;
8) the standard deviation of HT values;
9) the number of times the difference between ST and HT radius changes from negative to positive, or vice-versa.

IV. Simulations and Results

In this section, we presented the experimental results in order to access the quality of the proposed dataset and the feature extraction approach[5]. We carried out the experiments using three supervised pattern classifiers, namely Naïve Bayes, Optimum-Path Forest, and Support Vector Machines with RBF kernel. We also performed a 10-fold cross-validation to compute the mean recognition rates and standard deviation, being 90% of the dataset used for training and the remaining 10% for testing purposes[6].

The effectiveness of each classifier was evaluated according to four classification measures: Recall, Precision, F-score and Accuracy. Bellow, we briefly describe these four measures:

- Recall: is the true positive rate, being defined as the ratio $\frac{tp}{tp+fn}$, where tp and fn stand for the number of true positives and false negatives, respectively;
- Precision: is the true positive relevance rate, being defined as the ratio $\frac{tp}{tp+fp}$, where fp denotes the number of false positives;
- F-score: is the harmonic mean between Precision and Recall, being defined as the ratio $\frac{2tp}{2tp+fp+fn}$;
- Accuracy: recognition rate proposed by Papa et al. [8] that considers imbalanced datasets:

$$Acc = 1 - \frac{1}{2c}\sum_{i=1}^{c}\left(\frac{fp_i}{\sum\limits_{j\neq i, j=1}^{c} tp_j + fn_j} + \frac{fn_i}{tp_i + fn_i}\right), \qquad (7)$$

where tp_i, fp_i and fn_i denote the number of true positives, false positives and false negatives of class $i = 1, 2, \ldots, c$, respectively[7].

Table IV displays the experimental results for each classifier. As one can observe, the proposed dataset provided relevant rates according to the overall accuracy, mainly with respect to NB classifier, which achieved the highest accuracy rate, followed closely by OPF. Broadly speaking, these results may sound as a good indicator that the dataset has a potential to be used. However, as expected, the classifier's performance on the control group (i.e., the smallest class) was relatively poor when compared to the patient group. The only exception concerns with NB classifier, which obtained a better Recall rate over the control group. Since NB assumes the classes follow a normal distribution and its lower error bound is minimum under such assumption, it seems the control class fits better to a normal distribution than the patient group. Another interesting point concerns with the recall measure for both techniques considering control group: basically, we can observe the low values for such measure indicate that a considerable amount of control individuals have been misclassified as patient individuals, since this latter group has a high variability (a number of PD individuals have good handwritten skills so far).

Therefore, such results lead us to future works concerning ensemble of classifiers, in which NB may have a higher priority when labeling a given sample as belonging to the control group. We shall explore the good recognition rates for patient individuals obtained by OPF, as well as the very good accuracy obtained by NB over control data.

		NB	OPF	SVM
	Accuracy (%)	**78.9±3.5**	**77.1±3.8**	**75.8±3.3**
Control	F-score	0.38±0.04	0.30±0.10	0.23±0.11
	Precision	0.24±0.03	0.28±0.09	0.25±0.12
	Recall	0.91±0.09	0.32±0.12	0.24±0.14
Patient	F-score	0.49±0.10	0.82±0.03	0.83±0.03
	Precision	0.94±0.05	0.83±0.03	0.82±0.02
	Recall	0.34±0.10	0.80±0.05	0.83±0.06

Table I
Experimental results considering each classifier and accuracy measure.

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

[5]The proposed dataset and the extracted features are available at http://wwwp.fc.unesp.br/~papa/pub/datasets/Handpd

[6]SVM parameters have been optimized through a grid-search procedure within the ranges $C \in [2^{-5}, 2^{-3}, \ldots, 2^{13}, 2^{15}]$ and $\sigma \in [2^{-15}, 2^{-13}, \ldots, 2^{1}, 2^{3}]$, in which C and σ stand for SVM soft margin parameter and RBF kernel variance value, respectively.

[7]In our case, $c = 2$ since the proposed dataset has only two classes, i.e., patient and control individual.

V. Conclusions

In this paper, we dealt with the problem of Parkinson's Disease recognition by means of machine learning and computer vision techniques, being the main contributions to design a new dataset acquired by low-cost devices, as well as to propose a new quantitative feature to measure the "amount of tremor" of an individual.

The experimental setup was conducted over a 10-fold cross-validation procedure, using 90% of the dataset for training, and the remaining data for testing purposes. The accuracy of the proposed dataset has been assessed by means of a Naïve Bayes classifier, as well as by OPF and SVM techniques. The best result was obtained by NB classifier, which achieved about 78.9% of recognition rates, followed closely by OPF. Such accuracy rate appears to be very suitable for PD recognition, since we have employed a low-cost device for image acquisition. In regard to future works, we aim at considering meander images for PD identification, as well as to employ both meander and spiral images for decision-making-fusion purposes.

Acknowledgments

The authors would like to thank FAPESP grants #2013/20387-7 and #2014/16250-9, Capes and CNPq grants #306166/2014-3, #470571/2013-6 and #303182/2011-3.

References

[1] R. E. Burke, "Evaluation of the braak staging scheme for parkinson's disease: Introduction to a panel presentation," *Movement Disorders*, vol. 25, no. S1, pp. S76–S77, 2010.

[2] J. Parkinson, *An Essay on the Shaking Palsy*, Whittingham and Rowland, 1817.

[3] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun, "Collection and analysis of a parkinson speech dataset with multiple types of sound recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 828–834, July 2013.

[4] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1015–1022, April 2009.

[5] A. N. Montagnoli A. R. Gatto A. A. Spadotto J. C. Pereira, A. O. Schelp and L. R. Carvalho, "Residual signal auto-correlation to evaluate speech in Parkinson's disease patients," *Arquivos de Neuro-Psiquiatria*, vol. 64, pp. 912 – 915, 12 2006.

[6] M. A. Little, M. E. Patrick, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *BioMedical Engineering OnLine*, vol. 6, no. 1, pp. 1–19, 2007.

[7] T. E. Eichhorn, T. Gasser, N. Mai, C. Marquardt, G. Arnold, J. Schwarz, and W. H. Oertel, "Computational analysis of open loop handwriting movements in parkinson's disease: A rapid method to detect dopamimetic effects," *Movement Disorders*, vol. 11, no. 3, pp. 289–297, 1996.

[8] A. A. Spadotto, R. C. Guido, J. P. Papa, and A. X. Falcão, "Parkinson's disease identification through optimum-path forest," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2010, pp. 6087–6090.

[9] J. P. Papa, A. X. Falcão, and C. T. N. Suzuki, "Supervised pattern classification based on optimum-path forest," *International Journal of Imaging Systems and Technology*, vol. 19, no. 2, pp. 120–131, 2009.

[10] J. P. Papa, A. X. Falcão, V. H. C. Albuquerque, and J. M. R. S. Tavares, "Efficient supervised optimum-path forest classification for large datasets," *Pattern Recognition*, vol. 45, no. 1, pp. 512–520, 2012.

[11] F. S. Gharehchopogh and P. Mohammadi, "Article: A case study of parkinsons disease diagnosis using artificial neural networks," *International Journal of Computer Applications*, vol. 73, no. 19, pp. 1–6, July 2013, Published by Foundation of Computer Science, New York, USA.

[12] S. Pan, S. Iplikci, K. Warwick, and T. Z. Aziz, "Parkinsons disease tremor classification, a comparison between support vector machines and neural networks," *Expert Systems with Applications*, vol. 39, no. 12, pp. 10764 – 10771, 2012.

[13] T. Liu W. L. Zuo, Z. Y. Wang and H. L. Chen, "Effective detection of parkinson's disease using an adaptive fuzzy k-nearest neighbor approach," *Biomedical Signal Processing and Control*, vol. 8, no. 4, pp. 364 – 373, 2013.

[14] H. L. Chen C. Ma, J. Ouyang and X. H. Zhao, "An efficient diagnosis system for parkinson's disease using kernel-based extreme learning machine with subtractive clustering features weighting approach," *Computational and Mathematical Methods in Medicine*, vol. 2014, pp. 14, 2014.

[15] S. Haller, S. Badoud, D. Nguyen, V. Garibotto, K. O. Lovblad, and P. R. Roberthard, "Individual detection of patients with parkinson disease using support vector machine analysis of diffusion tensor imaging data: initial results," *American Journal of Neuroradiology*, vol. 33, no. 11, pp. 2123–2128, 2012.

[16] T. Y. Zhang and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Communications of the ACM*, vol. 27, no. 3, pp. 236–239, Mar. 1984.