

Parkinson's Disease Diagnosis using Spiral Test on Digital Tablets

Najd Al-Yousef¹, Raghad Al- Saikhan², Reema Al- Gowaifly³
Reem Al-Abdullatif⁴, Felwa Al-Mutairi⁵, Ouiem Bchir⁶

College of Computer and Information Sciences
Computer Science Department, King Saud University
Riyadh, Saudi Arabia

Abstract—For a proper diagnosis, Parkinson's disease (PD) requires frequent visits to the doctor for physical tests, causing a huge burden on the patient. As PD impairs the handwriting ability, the handwriting pattern can be used as an indicator for PD diagnosis. More specifically, the Static Spiral Test (SST) and the Dynamic Spiral Test (DST), that consists in retracing spirals using digital pen. Such exam can be self-conducted by the patient, and thus it would be convenient and non-time-consuming for both the patient and the medical staff. In this project, we designed and implemented a system that automatically self-aid-diagnoses PD using SST and DST on digital tablets. The system includes two main components, image processing techniques to pre-process and extract the appropriate visual features and machine learning techniques to recognize PD automatically. The conducted experiment showed that the semi-local Edge Histogram Descriptor extracted from DST drawing, and conveyed to a Gaussian Kernel Support Vector Machine outperforms the other considered systems with an accuracy, specificity and sensitivity around 90%.

Keywords—Component; Parkinson's disease (PD); computer-aided diagnosis; pattern recognition

I. INTRODUCTION

Parkinson's disease (PD) is a disorder that degenerates neurons which yields to failure of motor function because of smaller ratio of dopamine that is produced in brain [1], it affects the patient motor abilities such as speaking, writing, and walking. It has been estimated that around 7 to 10 million people worldwide have PD [2]. Its symptoms often appear gradually without being noticed by the patient. They are classified into those affecting movement (motor symptoms) and those that do not (non-motor symptoms).

Motor ones are easier to detect. One of the primary motor symptoms is the tremor which is an unintentional, rhythmic, slow muscle movement [3]. It occurs when the person is motionless and begins either in one hand, one foot, or one leg [2].

Even though PD cannot be cured, frequent monitoring is of high importance in order to obtain the proper diagnosis and help controlling the symptoms. Diagnosing Parkinson disease (PD) requires running different physical tests on the patient. These tests are performed by the physician in a clinic. Moreover, more than one visit is necessary to follow up with the patient's condition.

Given that PD is an age-related disorder where most patients are over 50 years old [4], who may have other age-related conditions, visiting the clinic may be an inconvenient task for them. Performing the tests remotely and comfortably at home, would decrease the burden of these tests and thus, encourage patients to do them. Moreover, that would decrease the load on the medical staff.

One solution to this problem could be to ask a patient to draw a certain pattern. Then, using image processing techniques extract the visual content of the drawn image and convey it to a machine learning algorithm that decides on the patient diagnosis.

Since PD is a type of movement disorder that impairs handwriting ability, we propose a system that will detect PD using the handwriting pattern. We will use both the traditional Static Spiral Test (SST) and the Dynamic Spiral Test (DST). SST requires the patients to follow a static spiral drawing. For the DST, the spiral that the patient is supposed to follow appears and disappears during the test time, so that the patient has to memorize the pattern and keep drawing it [5].

The proposed approach is intended to diagnose Parkinson's disease using both the static and the dynamic spiral tests independently and jointly. It will use image processing techniques to extract the needed feature from the spiral images and machine learning to automatically recognize Parkinson's disease.

For the feature extraction step, we intend to consider the following visual descriptors:

- Histogram of oriented gradients (HOG) [6],
- Edge Histogram Descriptor (EHD) [6],
- An application designed feature based on acceleration as described in [5], and
- Deep learning Auto-encoder generated feature.

For the classification step, we use Support Vector Machine (SVM) [7].

This paper is organized as follows: Section II presents and discusses the related works. Section III describes the methodology adopted in this research. Section IV sets the experiments that are conducted to assess the performance of the proposed system. Section V. reports the obtained

experimental results and analyze them. Lastly, Section VI concludes and summarizes the paper.

II. RELATED WORKS

In the following we outline the computer-based approaches that are related to our work. They are either based on gait, voice or handwriting pattern.

A. Gait based Approaches

Previous studies have shown that how the person walk can be a prediction of developing PD disease. In fact, Bridenbaugh and Kressig (2013) [8] concluded that there is an association between gait and cognition and that elderly people with gait impairments were more likely to develop cognitive impairments and problems in memory in addition to weaknesses in some processing functions.

A gait-based approach to detect PD disease is proposed in [9]. The step length can be estimated using the change in the waist height and the leg angle while walking. In order to record the gait characteristics such as step frequency and stride length, the Pedestrian Dead Reckoning (PDR) system and the smartphones' accelerometer sensor are used. The classification is done through SVM classifier [7].

B. Voice based Approaches

Recent research has been conducted to study the connection between PD and speech impairment such as dysphonia [10]. The authors in [11], introduced an algorithm for PD diagnosis based on voice analysis. The test consists of pronouncing the letter "A" for 3 seconds. From the recorded sound; 22 features are extracted. They are based on the pitch, the jitter, the shimmer and the noise ratio. Then, feature selection is performed using genetic algorithm (GA) [12]. Afterwards, SVM classifier [7] is used to distinguish between PD and healthy subjects.

The authors in [13], developed a clinical expert system to detect PD from the subject's vocals. The system extracts three voice recordings of each subject pronouncing the letter "A" for 5 seconds and then uses a waveform matching algorithm [14] to extract 44 acoustic features which are based on: noise, pitch perturbation, amplitude of the spectral envelope measures and nonlinear ones. Then, it uses the Bayesian classifier [15] to distinguish PD patients from non-PD ones.

C. Handwriting based Approaches

Drotár et al. (2016) [16] introduced a system of aided Parkinson's diagnosis based on kinematic characteristics and pressure in handwriting. The dataset of the proposed approach includes the task of drawing an Archimedean spiral. The entropy feature based on the pressure applied on the writing surface is extracted, then the SVM classifier is applied for the recognition task [7].

The authors in [17], suggested an approach to diagnose PD by using the handwriting pattern of the patient by conducting two tests, the spiral and the meander tests. The spiral test requires the subject to draw a spiral while the meander requires the patient to draw more structural lines. The structural Cooccurrence Matrix feature is extracted from both drawings. The features of the spiral test and those from

the meander test are conveyed separately to the classifier and also combined. Then three classifiers are used separately. They are Naïve Bayes [18], Optimum-Path Forrest OPF [19], and Support Vector Machine SVM [7]. The results showed that the highest accuracy rate is obtained when using spiral test and SVM classifier [7].

The authors in [5], proposed a static and dynamic spiral tests diagnosis approach for PD that uses the dynamic spiral test and the static spiral test referred to as DST and SST respectively. These two tests are conducted using an electronic equipment like a computer or a tablet.

SST consists in providing the subject with a static spiral test, the subject is then required to follow the pattern in order to draw a spiral. DST consists in providing the patient with the dynamic spiral. In other words, the spiral appears and disappears periodically. The subject has to continue drawing the pattern even if the model disappears. From both the DST and the SST spiral, the acceleration feature is extracted and then the dissimilarity of acceleration histograms (DAH) is computed. DAH is then used as an indicator, such that a small DAH means that the subject is healthy while a large one means he is a PD patient. We should mention here that the authors in [5] did not use any machine learning techniques to recognize PD patients. Moreover, they did not suggest a way of specifying a threshold allowing to discriminate between healthy and PD subjects.

III. METHODOLOGY

In this project, we proposed and designed a pattern recognition system that is able to analyze the obtained results automatically. The main components of this system are:

- 1) Feature extraction, which translates the visual content of the drawn spiral image into a numerical vector.
- 2) Feature combination which allows to combine SST and DST tests.
- 3) Machine learning to learn a model that allows discrimination between the drawings of a PD patient and non-PD patient, and thus yields the categorization of the patients as PD or non-PD.

From the SST and DST drawn images, we extract four features which are:

- Histogram of Oriented Gradients [20],
- Edge Histogram Descriptor [21],
- Acceleration based descriptor [5],
- Auto-encoder generated descriptor [22].

In the following, we give a brief description of each component of the system and motivate its choice.

A. Histogram of Oriented Gradients

Histogram of Oriented Gradients (HOG) is a feature descriptor that is being widely used in many applications [20], [23]. It describes the edges present in an image by using the gradient at each pixel. In fact, the gradient is a mathematical

tool that allows measuring the direction and amplitude of the change in the pixels' values.

The gradient of an image at pixel $F(x,y)$ is defined as:

$$\nabla F = \left[\frac{\delta F}{\delta x}, \frac{\delta F}{\delta y} \right]$$

Where, ∇F is the gradient of the pixel, $\frac{\delta F}{\delta x}$ is the partial derivative of the pixel with respect to x, and $\frac{\delta F}{\delta y}$ is the partial derivative of the pixel with respect to y.

The gradient direction [24], that reflects the direction of the most rapid increase in the pixel intensity, can be expressed as:

$$\theta = \tan^{-1} \left[\frac{\delta F}{\delta y} / \frac{\delta F}{\delta x} \right]$$

The gradient magnitude reflects the edge strength [24], and it can be expressed as:

$$\|\nabla F\| = \sqrt{\left(\frac{\delta F}{\delta x}\right)^2 + \left(\frac{\delta F}{\delta y}\right)^2}$$

The gradient descriptor is computed using the gradient filter. Each pixel and its neighborhood pixels are correlated using the filter in Fig. 1.

The following steps describe the HOG algorithm:

- 1) Compute the gradient vector using the filter in Fig. 4.
- 2) Compute the 9-bin histogram of the obtained direction (20° is used for each direction) [20].
- 3) Compute the 9-bin histogram of the obtained amplitudes.
- 4) The HOG vector is then the concatenation of the sub histograms obtained in steps 3 and 4.

HOG describes the edges in an image by computing the gradient at each pixel. This can be achieved by using the HOG filter [24]. The obtained gradient magnitude and direction of each pixel are used to compute two 9-bin histograms.

For both SST and DST, the main difference between the spirals drawn by a PD patient and a non-PD patient is the shape of the spiral. Our main interest is the edges' oscillations of the drawn spirals. Since HOG descriptor reflects the edges in the image it is believed to discriminate between the structure/shape of the spiral between PD and non-PD patient.

0	0	0
$-\frac{1}{2}$	0	$\frac{1}{2}$
0	0	0

Fig 1. The Gradient Filter.

B. Edge Histogram Descriptor

Edge Histogram Descriptor (EHD) uses histograms to describe local and global edge distribution [21]. It represents the shape content of an image. The edge of each pixel is classified into five types: vertical, horizontal, 45-degree diagonal, 135-degree diagonal, and non-directional edges. The pixel edge type is determined by correlating each pixel and its neighboring pixels by edge detection filters. Fig. 2 shows the five filters corresponding to the 5 considered directions.

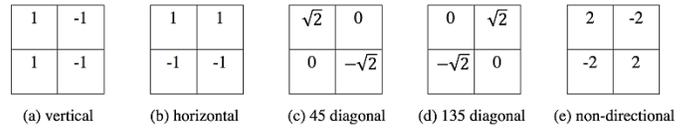


Fig 2. Edge Filters.

After filtering the image with the 5 filters, the 5 results are compared. Each pixel is assigned the direction of the filter type that has the maximum response. Then, three types of histograms are computed: global, local and semi-local histogram.

The global histogram is computed over the whole image and thus it is a 5-length vector, one entry for each direction. For the local histogram, the original image is partitioned into 4*4 non-overlapping blocks called sub-image which yields a total of 16 sub-images as shown in Fig. 3.

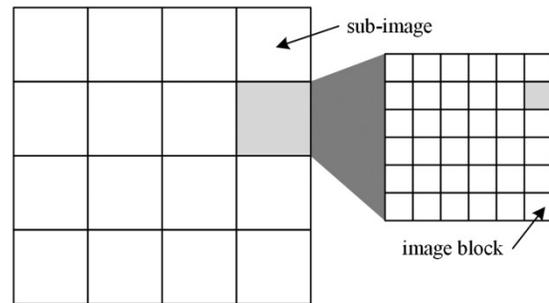


Fig 3. Image Partitioning into 16 blocks [13].

For each sub-image an edge histogram is generated to represent the frequency of occurrence of the different types of edges in each image-block. Generating 5-bin histogram for each of the 16 sub-images yields a total of 80 bin histogram.

The semi-local histogram is obtained by concatenating 13 histograms as follows:

- 4 Histograms are computed by summing the local histograms over each row of blocks (Fig. 7(a)).
- 4 other histograms are generated by summing the local histograms over each column of blocks (Fig. 7(b)).
- One histogram is computed by summing the 4 upper-left sub-blocks' local histograms (Fig. 7(c) block A).
- One histogram is obtained by summing the 4 upper-right sub-blocks' local histograms (Fig. 7(c) block B).
- One histogram is generated by summing the 4 down-left sub-blocks' local histograms (Fig. 7(c) block C).

- One histogram is generated by summing the 4 down-right sub-blocks' local histograms (Fig. 7(c) block D).
- One histogram is generated by summing the 4 central sub-blocks' local histograms (Fig. 7(c) block E).

Fig. 4 displays how the image is partitioned in order to compute the semi-local histograms.

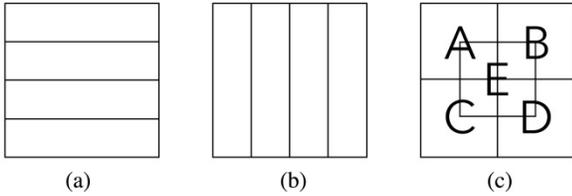


Fig 4. Image Partitioning for Semi-Local Histogram Computation. (a) Into 4 Rows (b) Into 4 Columns (c) Into 4 Sub-Block Sets.

Edge histogram descriptor (EHD) is one of the most common methods for detecting the shape [21]. It describes the global, semi-local and local edges in an image. In fact, the image is partitioned into image blocks. For each image block an edge distribution histogram is computed by categorizing the edges into five types using edge detection filters.

Since EHD captures both local and global edges, it would be able to discriminate between the shape of PD and non-PD spirals through an appropriate description of the edge oscillations.

C. Acceleration

The authors in [5] propose an application dedicated feature for SST and DST drawing where at each time t , the coordinates of the drawing (x_t, y_t) are recorded. The velocity is then defined as the distance between two consecutive samples at $t-1$ and t as expressed as follows:

$$V_t = \sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2}$$

Since PD patients are more likely to be confused leading to instantaneous speedup or slowdown, the velocity changes during the drawing for both tests. In other words, instantaneous acceleration changes at time t . This acceleration is defined as the difference between the velocity at time (t) and at time $(t-1)$. Expressed as in:

$$A_t = V_t - V_{t-1}$$

Acceleration histograms are then computed for SST and DST drawing. Then, the dissimilarity between the two obtained histograms, DAH, is computed as:

$$DAH = \|H_{SST} - H_{DST}\| = \sum_{i=1}^{10} (H_{SST}(i) - H_{DST}(i))^2$$

People who don't suffer from PD are expected to show similar performances in SST and DST and thus would get a DAH close to zero.

We should mention here that the acceleration based on SST and DST test could be used separately as a feature descriptor. It can also be used to combine both tests as described in the previous formula.

Since PD patients are expected to be confused during the drawing of SST and DST leading to instantaneous speedup or slowdown, the velocity and acceleration change for both tests. This feature is used by the authors in [5] suggesting the use of SST and DST.

D. Auto-Encoder

An auto-encoder is an unsupervised neural network learning algorithm. Auto-encoders take unlabeled data as an input, encode them and then try to reconstruct the encoded data as accurately as possible. An auto-encoder is composed of an input layer, an output layer and several hidden layers as shown in Fig. 5.

In order to compute the neural-net weights for each hidden layer, the previous layer is considered as an input layer and as an output layer. This yields a 3-layer network as shown in Fig. 6.

Then, the network parameters are determined as for any standard 3-layer artificial neural-net. This process is repeated for all the hidden layers starting from the first one where the input layer is the actual input data.

Once the network can accurately reconstruct the input this means that the hidden layer contains enough information to represent the output. Thus, an auto-encoder may act as a feature extraction engine by using the encoding layer as a feature descriptor [22].

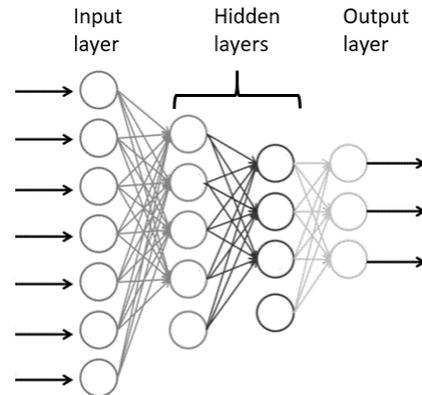


Fig 5. Layers of an Auto-Encoder.

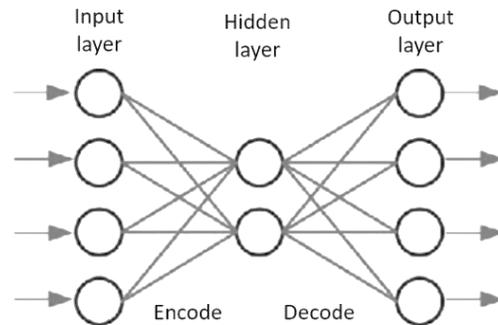


Fig 6. A Three-Layer Auto-Encoder.

An auto-encoder is an unsupervised neural network with multiple hidden layers that is trained with a standard weight

adjustment algorithm to reproduce the input with fewer unites which compels the hidden layer units to become feature detectors and then predict the classes based on the output from the previous layers. As auto-encoder have shown great results for knowledge extraction [22], it would be able to capture the intrinsic characteristic of the drawings and thus, it would be able to distinguish PD and non-PD patients.

E. Combinations of SST and DST Features

Consider SST and DST drawings of the same patient j . Let $F_{ij}^{(1)}$ be the feature i extracted from the SST image of patient j and let $F_{ij}^{(2)}$ be the feature i extracted from the DST image of patient j . We consider to approaches to combine $F_{ij}^{(1)}$ and $F_{ij}^{(2)}$ into a single vector F_{ij}^c .

1) *Concatenation*: In order to obtain the combined feature F_{ij}^c for patient j with respect to feature i , we concatenate $F_{ij}^{(1)}$ and $F_{ij}^{(2)}$ as follows:

$$F_{ij}^c = [F_{ij}^{(1)} \ F_{ij}^{(2)}]$$

The dimensionality of F_{ij}^c will be the double of the dimensionality of feature i .

2) *Difference*: We can combine $F_{ij}^{(1)}$ and $F_{ij}^{(2)}$ by considering their difference. We define F_{ij}^c as:

$$F_{ij}^c = \sum_{k=1}^d (F_{ijk}^{(1)} - F_{ijk}^{(2)})^2$$

Where d is the dimensionality of feature i , $F_{ijk}^{(1)}$ is the k^{th} entry of feature vector $F_{ij}^{(1)}$, and $F_{ijk}^{(2)}$ is the k^{th} entry of feature vector $F_{ij}^{(2)}$.

When using the concatenation approach, we assume that both the information from SST and DST are useful and we will convey them both at the same time to the classifier.

On the other hand, using the difference as a method of combining the two tests, assumes that the performance of a PD patient will be better for SST test than for DST test since DST involves testing the short memory of the patient, and thus the difference would be meaningful as a way of discrimination.

F. Support Vector Machine

A Support Vector Machine is a supervised machine learning technique for the purpose of binary data classification. It finds the optimal hyperplane that separates two classes. As reported, most of previous works used SVM [7]. In fact, it has been proven to be an affective classifier for 2-class problems [7].

IV. EXPERIMENTAL DESIGN

A. Dataset Description

The dataset used to assess the proposed system is obtained from [5] [25]. It was collected from 15 non-PD patients and 57 PD patients. Knowing that both SST and DST were used, that

gives a total of 30 non-PD and 114 PD recordings. A sample from the dataset is displayed in Fig. 7.

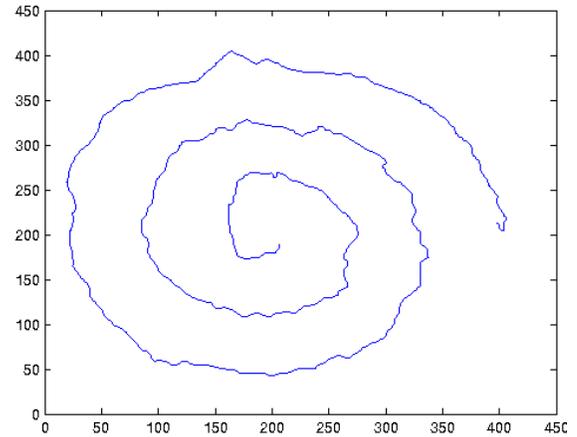


Fig 7. A Sample from the Dataset.

The drawings are recorded using a digitized graphics tablet shown in Fig. 8.



Fig 8. Digitized Graphics Tablet [5].

B. Assessment Technique

As an assessment technique, we use the K-fold-cross-validation [26] -with k equal to 10-. It divides the training set randomly into K subsets of relatively the same size. The K^{th} subset is used for testing while the first $K-1$ subsets are used as the training set. This process is repeated K times so that each subset is used as the testing set once. Then the performance of the cross-validation is defined as the average performance of the K iterations [27].

C. Assessment Measures

To assess the proposed approach, we use different statistics. Specifically, accuracy, sensitivity and specificity.

- Accuracy [28], measures the overall correctness of the classifier:
- Sensitivity [28], (also known as Recall or True Positive Rate) measures the correctly classified positive tests over the actual positive.
- Specificity [28] measures the correctly classified negative tests over the actual negative.

We should mention here that since the data is unbalanced (57 PD patients and 15 control), the accuracy maybe misleading. On the other hand, the sensitivity is more appropriate, since it is more important to detect wrongly a PD patient than missing one.

In addition to the previously mentioned statistical performance measures, The Receiver Operating Characteristic (ROC) [29] is also a performance technique that is used to assess the recognition system over all possible thresholds. ROC curves of all the possible approaches are plotted by sketching the 1-Specificity on the x-axis against the Sensitivity on the y-axis.

Area Under the Curve (AUC) is the total area underneath the ROC curve. It measures the performance of a recognition system. If AUC is 1.0, it means that the system can perfectly distinguish between two classes.

D. Experiment Description

The first step of the experiment consists in extracting the 4 considered features from SST and DST drawings. This yields 19 features matrices:

- $\Pi 1$: HOG feature from SST
- $\Pi 2, \Pi 3, \Pi 4, \Pi 5$: EHD features from SST. We will consider the local EHD, semi-local EHD, the global EHD, and the overall EHD descriptors separately.
- $\Pi 6$: acceleration feature from SST
- $\Pi 7, \Pi 8, \Pi 9$: Auto-encoder features from SST. We consider 6 layers with different number of nodes which are 1000, 500, 100, 64, 32, and 16 as shown in Fig. 9. As input to the SVM classifier, we convey the output of the layers of size 64, 32, and 16 separately.
- $\Pi 10$: HOG feature from DST
- $\Pi 11, \Pi 12, \Pi 13, \Pi 14$: EHD feature from DST. We will consider the local EHD, semi-local EHD, the global EHD, and the overall EHD descriptors separately.
- $\Pi 15$: acceleration feature from DST
- $\Pi 16, \Pi 17, \Pi 18, \Pi 19$: Auto-encoder features from DST. We consider 6 layers with different number of nodes which are 1000, 500, 100, 64, 32, and 16 as shown in Fig. 9. As input to the SVM classifier, we convey the output of the layers of size 64, 32, and 16 separately.

In the second step we combine SST and DST using the concatenation approach. This gives 9 feature matrices:

- $\Pi 20$: HOG from concatenating SST and DST
- $\Pi 21$: Local EHD from concatenating SST and DST
- $\Pi 22$: Semi-local EHD from concatenating SST and DST
- $\Pi 23$: Global EHD from concatenating SST and DST
- $\Pi 24$: Overall EHD from concatenating SST and DST
- $\Pi 25$: acceleration from concatenating SST and DST

- $\Pi 26$: auto-encoder feature of size 64 from concatenating SST and DST
- $\Pi 27$: auto-encoder feature of size 32 from concatenating SST and DST
- $\Pi 28$: auto-encoder feature of size 16 from concatenating SST and DST

Then, we combine SST and DST tests using the difference approach. We get 9 feature matrices:

- $\Pi 29$: HOG from the difference between SST and DST
- $\Pi 30$: Local EHD from the difference between SST and DST
- $\Pi 31$: Semi-local EHD from the difference between SST and DST
- $\Pi 32$: Global EHD from the difference between SST and DST
- $\Pi 33$: Overall EHD from the difference between SST and DST
- $\Pi 34$: acceleration from the difference between SST and DST
- $\Pi 35$: auto-encoder feature of size 64 from the difference between SST and DST
- $\Pi 36$: auto-encoder feature of size 32 from the difference between SST and DST
- $\Pi 37$: auto-encoder feature of size 16 from the difference between SST and DST

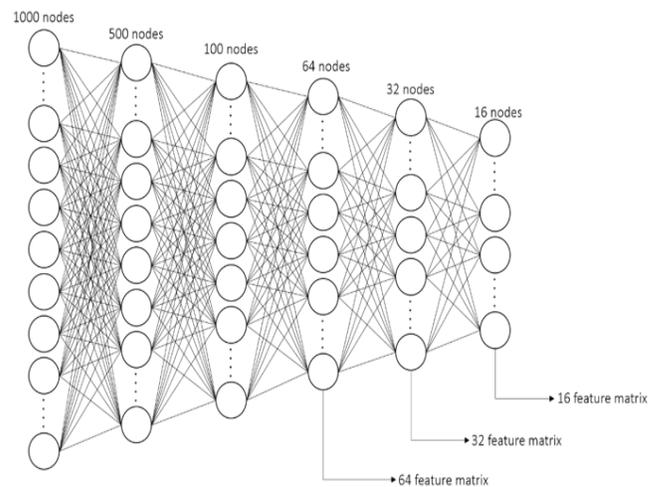


Fig 9. Number of Nodes in each Layer of the Encoding Part in the Auto-Encoder.

Finally, each of the obtained 37 matrices ($\Pi 1 \dots \Pi 37$) will be conveyed separately to an SVM classifier.

Besides, we will consider three variants of the SVM classifier. More specifically, in the first experiment, we assume that the data is well separated and the boundary between the two classes is linear, and thus we use hard-margin SVM [30]. In the second experiment, we assume that the data

is not well separated and that an over-fitting problem may occur and conduct the experiment on a soft-margin SVM [31] to check if this assumption holds. In the third experiment, we assume that the boundary separating the two classes is not linear and that a mapping of the extracted features to a new feature space is necessary. Therefore, we use the Gaussian kernel-SVM [32]. We should mention here that the 37 extracted features are considered in each of the three experiments and conveyed separately to each considered classifier.

Lastly, the performance measures described previously are computed for each considered experiment with respect to the 37 features. The analysis and comparison of the results would allow to conclude on the design of the PD diagnosis system and its effectiveness.

V. RESULTS AND DISCUSSION

To explore the possibility that the data is not linearly separable, we use Gaussian kernel SVM [32]. The Gaussian kernel happens to be one of the most important algorithms among the kernel based, it maps the data features onto a higher dimensional space using the Gaussian kernel function [32].

Fig. 10-14 show the performance results of using HOG, EHD, Acceleration and Auto-encoder features as input to Gaussian kernel SVM [32].

Fig. 10 displays the performance results when using the HOG feature on SST, on DST, on the concatenation of both tests, and on the difference between them. The best performance is obtained when extracting HOG from DST drawings with a sensitivity of 85.67% and specificity of 45%.

Fig. 11 shows the performance results of the global, semi-local, local and the overall EHD feature on SST, on DST, on the concatenation of both tests, and on the difference between them. As we can see, the semi-local EHD extracted from DST drawings gives the best results which is 91% for sensitivity and 90% for specificity.

Fig. 12 shows the performance results when using Acceleration feature as input for the Gaussian Kernel SVM [32]. We notice that concatenation of the acceleration feature of SST and DST achieves the best results with a sensitivity of to 98.3% and specificity of 25%.

Fig. 13 displays the accuracy, sensitivity and specificity results when conveying the auto-encoder feature as input to the Gaussian kernel SVM [32]. We notice that the concatenation of the 32 encoded layers of SST and DST gives the best results with a sensitivity and specificity of 95% and 10%, respectively.

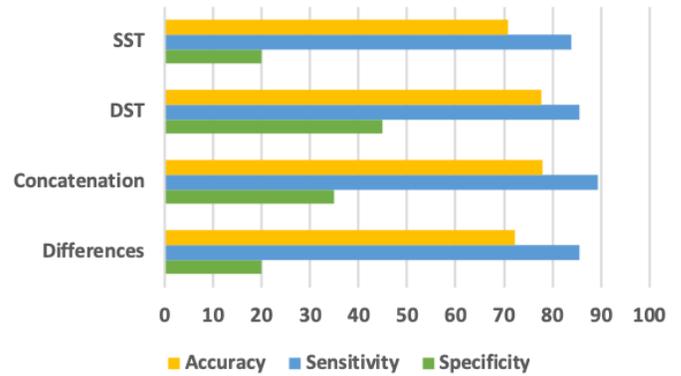


Fig 10. HOG Descriptor Performance using Gaussian kernel SVM.

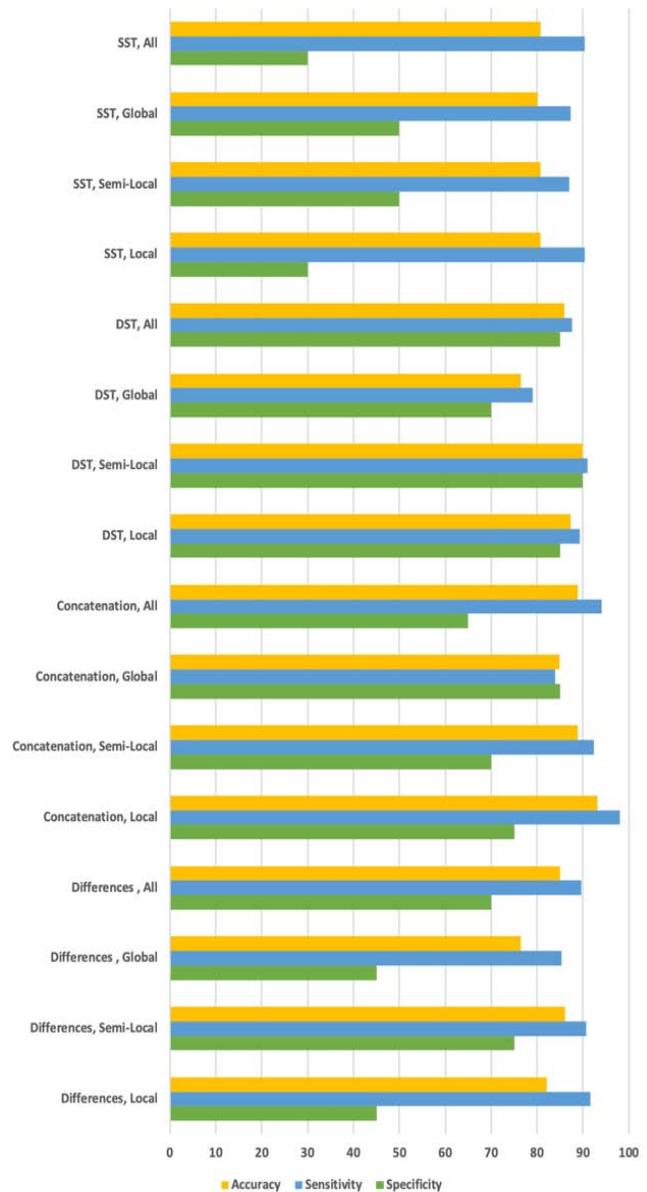


Fig 11. EHD Descriptor Performance using Gaussian kernel SVM.

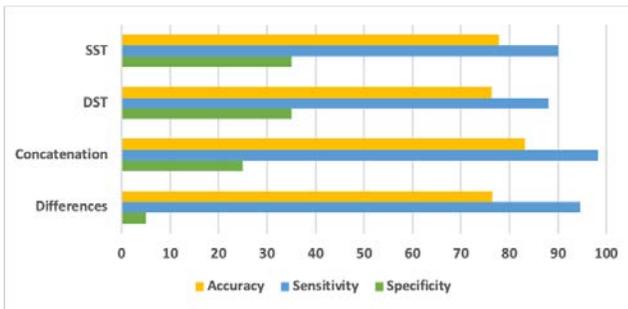


Fig 12. Acceleration Features Performance using Gaussian kernel SVM.

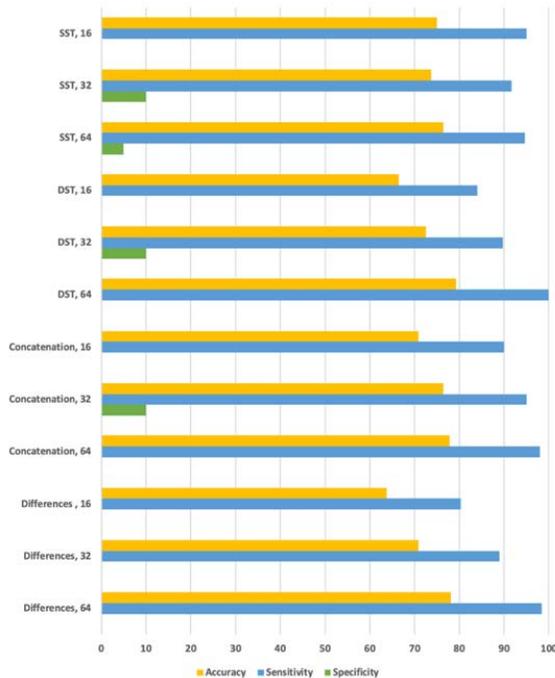


Fig 13. Auto-Encoder Features Performance using Gaussian kernel SVM.

Fig. 14 and Table I show the ROC and the AUC when using Gaussian Kernel SVM [32]. We can see that the semi-local EHD extracted from DST drawings outperforms the other features.

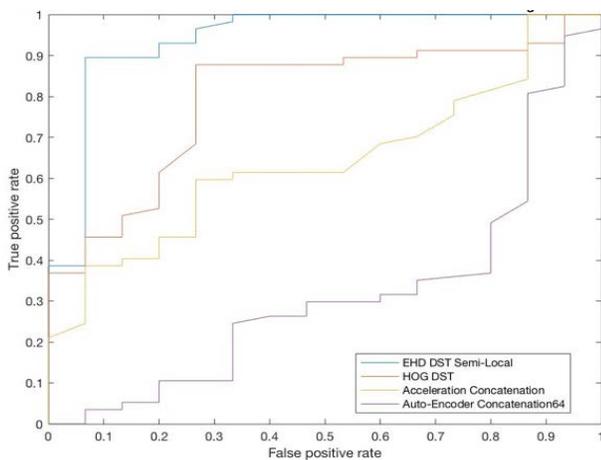


Fig 14. ROC Curves using Gaussian Kernel SVM.

TABLE I. SUMMARY OF THE BEST AUC VALUES USING GAUSSIAN KERNEL SVM

Best AUC Values	Experiment details		
	Size	Type of test	AUC
HOG	18	DST	0.797661
EHD	65	DST	0.938012
Acceleration	32	Concatenation	0.645029
Auto-encoder	128	Concatenation	0.314620

The results of semi-local EHD extracted from DST drawings outperforms all the other descriptors, when using Gaussian kernel SVM [32], it has an accuracy of 90%, sensitivity of 91.3%, and specificity of 90%. Gaussian kernel SVM [32] had an astounding improvement compared to hard and soft margin SVM [30] [31]. This indicates that the data is not linearly separable and mapping the data points using the Gaussian kernel SVM [32] allowed SVM to better classify the data.

In conclusion, the semi-local EHD descriptor extracted from DST drawing, and conveyed to a Gaussian kernel SVM [32] gives the best results. This is due to the fact that semi-local EHD describes 13 overlapping regions of the image separately and then concatenates the obtained histograms. This gives a good local edge information of the spiral drawings and allow distinguishing PD from non-PD. Moreover, the Gaussian kernel SVM [32] was able to classify the data better than hard-margin SVM [30] and soft-margin SVM [31] since it allowed mapping the data to a new feature space in which the data is linearly separable. Therefore, for the PD aided diagnosis system, we propose the system displayed in Fig. 15.

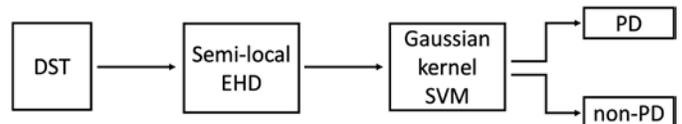


Fig 15. PD Aided Diagnosis System.

TABLE II. FEATURE EXTRACTION RUNNING TIME PER IMAGE

Feature extraction running time		Time (seconds)
HOG		3.72
EHD	All	7.82
	Global	4.47
	Local	4.05
	Semi-Local	4.54
Acceleration		9.3361
Auto-encoder	16	17.940
	32	17.914
	64	17.843

The proposed system contains two main components which are feature extraction and SVM classifier. Table II depicts the time need to extract each considered feature in seconds. We should mention here since the semi-local EHD descriptor gives the best results, the feature extraction phase will take only 4.54 seconds for the obtained system. Moreover, the time complexity for running SVM is linear with respect to the number of support vectors.

VI. CONCLUSION

Recently, the number of people with PD has augmented considerably. This makes it one of the major health problems.

Since it has no cure, an early detection is very important in order to allow an appropriate treatment. Moreover, it is crucial to monitor regularly the progress of the symptoms. However, this requires the patient to often visit the physician dealing with transportation, waiting, appointments, etc. This is inconvenient, especially that PD affects mostly elderly people. Besides, it involves the physician time and efforts.

In this project, we alleviated the monitoring of PD by designing a self-conducted test that uses recent technology advances along with pattern recognition techniques.

As a typical pattern recognition system includes a feature extraction step and a classification step in this project, we described the features extraction techniques that we investigated. We also outlined the machine learning technique, SVM classifier that will be applied.

A review of computer-based PD detection approaches using new technologies was outlined in this work. These approaches are based either on image or signal data. The latter source of data concerns gait on voice pattern analysis while image data is related either to the analysis of brain images or the analysis of handwriting pattern.

During these experiments, we used SST and DST image data gathered from a tablet device [5]. We investigated several features and conveyed them to an SVM classifier. We also investigated each test separately, SST and DST, and two ways of combining them. We implemented and assessed their performances. After analyzing the results, we conclude that the semi-local EHD [21] extracted from DST drawing, and conveyed to a Gaussian Kernel SVM [32] outperforms the other considered systems with an accuracy, specificity and sensitivity around 90%.

In order to investigate further the use of deep learning in extracting the visual descriptors from SST and DST drawing and its ability to discriminate between PD and non-PD drawing pattern, we plan to collect more data. The large size of the data would allow an effective training of the deep learning network.

Although the Gaussian kernel SVM gave good results, it may be enhanced further. As future work, we plan to investigate the use of Gaussian Mixture Classifier (GMM) [33]. In fact, GMM can model each class with several Gaussian and thus, can deal with variability of the data within each class.

Another way of enhancing the results would be to use fusion techniques. They could be applied on a set of classifiers, a set of visual descriptors, or a set of different data sources. The latter could be done by using the drawing pattern along with the speech pattern to discriminate between PD and non-PD patients.

REFERENCES

- [1] S. N. Heyn, C. P. Davis and M. C. Stöppler, "Parkinson's Disease Symptoms, Causes, Stages, Treatment, and Life Expectancy," MedicineNet, 28 August 2018. [Online]. Available: https://www.medicinenet.com/parkinsons_disease/article.htm. [Accessed 8 September 2018].
- [2] "Tremor in Parkinson's," American Parkinson Disease Association, [Online]. Available: <https://www.apdaparkinson.org/what-is-parkinsons/symptoms/tremor/>. [Accessed 8 September 2018].
- [3] MedicineNet, "Tremors," MedicineNet, [Online]. Available: <https://www.medicinenet.com/tremor/article.htm>. [Accessed 8 September 2018].
- [4] Parkinson's News Today, "Parkinson's Disease Statistics," Parkinson's News Today, [Online]. Available: <https://parkinsonsnewstoday.com/parkinsons-disease-statistics/>. [Accessed 17 September 2018].
- [5] M. E. Isenkul, B. E. Sakar and O. Kursun, "Improved spiral test using digitized graphics tablet for monitoring Parkinson's disease," in The 2nd International Conference on e-Health and Telemedicine, Istanbul, 2014.
- [6] Sikora, Thomas, "The MPEG-7 visual standard for content description-an overview," IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 6, pp. 696-702, June 2001.
- [7] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer and D. Haussler, "Article Navigation Support vector machine classification and validation of cancer tissue samples using microarray expression data," Bioinformatics, vol. 16, no. 10, p. 906-914, October 2000.
- [8] S. Bridenbaugh and R. W. Kressig, "Quantitative Gait Disturbances in Older Adults with Cognitive Impairments," Current Pharmaceutical Design, vol. 20, no. 19, pp. 3165-3172, 2014.
- [9] K.-C. Lan and W.-Y. Shih, "Early Diagnosis of Parkinson's Disease using a Smartphone," in The 11th International Conference on Mobile Systems and Pervasive Computing (MobiSPC-2014), Tainan, 2014.
- [10] M. Little, P. Mcsharry, E. Hunter, J. Spielman and L. Ramig, "Suitability of Dysphonia Measurements for Telemonitoring of Parkinsons Disease," IEEE Transactions on Biomedical Engineering, vol. 56, no. 4, pp. 1015-1022, 2009.
- [11] M. Shahbakhi, D. T. Far and E. Tahami, "Speech Analysis for Diagnosis of Parkinson's Disease Using Genetic Algorithm and Support Vector Machine," Journal of Biomedical Science and Engineering, vol. 7, no. 4, pp. 147-156, 2014.
- [12] D. E. Goldberg, Genetic algorithms in search, optimization, and machine learning, vol. 27, New York: Addison-Wesley, 1989.
- [13] L. Naranjo, C. J. Pérez, Y. Campos-Roca and J. Martín, "Addressing voice recording replications for Parkinson's disease detection," Expert Systems with Applications, vol. 46, no. 15, pp. 286-292, 2016.
- [14] I. R. Titze and H. Liang, "Comparison of Fo Extraction Methods for High-Precision Voice Perturbation Measurements," Journal of Speech, Language, and Hearing Research, vol. 36, no. 6, pp. 1120-1133, 1993.
- [15] M. Singh and G. M. Provan, "Efficient learning of selective Bayesian network classifiers," in Thirteenth International Conference on Machine Learning, Bari, Italy, 1996.
- [16] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal and M. Faundez-Zannuy, "Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease," Artificial intelligence in medicine, vol. 67, no. 1, pp. 39-46, 2016.
- [17] J. W. M. de Souza, S. S. A. Alves, E. d. S. Rebouças, J. S. Almeida and P. P. R. Filho, "A New Approach to Diagnose Parkinson's Disease Using a Structural Cooccurrence Matrix for a Similarity Analysis," Computational Intelligence and Neuroscience, vol. 2018, pp. 1-8, 2018.

- [18] F. Keller, "Naive Bayes Classifiers," Saarland University, 3 March 2005. [Online]. Available: <http://www.coli.uni-saarland.de/~crocker/courses/learning/lecture10.pdf>. [Accessed 20 October 2018].
- [19] J. P. Papa, A. X. Falcão, V. H. C. de Albuquerque and J. M. R. Tavares, "Efficient supervised optimum-path forest classification for large datasets," *Pattern Recognition*, vol. 45, no. 1, pp. 512-520, 2012.
- [20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE*, San Diego, 2005.
- [21] C. S. Won, S.-J. Park and D. K. Park, "Efficient Use of MPEG-7 Edge Histogram Descriptor," *ETRI Journal*, vol. 24, no. 1, pp. 23-30, February 2002.
- [22] C.-Y. Liou, W.-C. Cheng, J.-W. Liou and D.-R. Liou, "Autoencoder for words," *Neurocomputing*, vol. 139, pp. 84-96, 2014.
- [23] P. H. Kassani and A. B. j. Teoh, "A new sparse model for traffic sign classification using soft histogram of oriented gradients," *Applied Soft Computing*, vol. 52, no. C, pp. 231-246, 2017.
- [24] Udacity, "Computational Photography by Georgia Tech," Udacity, 23 February 2015. [Online]. Available: <https://sa.udacity.com/course/computational-photography--ud955>. [Accessed 22 Septemser 2018].
- [25] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbaş, F. Gurgun, 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, 2013.
- [26] I. H. Witten, E. Frank, M. A. Hall and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, Cambridge: Elsevier, 2011.
- [27] K. Duan, S. S. Keerthi and A. N. Poo, "Evaluation of simple performance measures for tuning SVM hyperparameters," *Neurocomputing*, vol. 51, pp. 41-59, 2003.
- [28] C. E. Metz, "Basic principles of ROC analysis," *Seminars in Nuclear Medicine*, vol. 8, no. 4, pp. 283-298, 1978.
- [29] R. Kumar and A. Indrayan, "Receiver operating characteristic (ROC) curve for medical researchers," *Indian Pediatrics*, vol. 48, no. 4, pp. 277-287, 2011.
- [30] C.-W. Hsu, C.-C. Chang and C.-J. Lin, "A Practical Guide to Support Vector Classification," National Taiwan University, Department of Computer Science, Taipei, 2010.
- [31] J. Shawe-Taylor and N. Cristianini, "On the generalization of soft margin algorithms," *IEEE Transactions on Information Theory*, vol. 48, no. 10, pp. 2721 - 2735, 2002.
- [32] Z. Liu and H. Xu, "Kernel Parameter Selection for Support Vector Machine Classification," *Journal of Algorithms & Computational Technology*, vol. 8, no. 2, pp. 163-177, 2013.
- [33] A. M. Patrikar and J. P. Baker, "Improving accuracy of Gaussian mixture model classifiers with additional discriminative training," in *International Joint Conference on Neural Networks (IJCNN)*, Vancouver, 2016.