Parkinson's Disease Identification Through Deep Optimum-Path Forest Clustering

Luis C. S. Afonso*, Clayton R. Pereira*, Silke A. T. Weber†, Christian Hook‡ and João P. Papa§
*UFSCar - Federal University of São Carlos, Department of Computing, São Carlos, Brazil
Email: {sugi.luis, clayton.pereira}@dc.ufscar.br
†UNESP - São Paulo State University, Medical School, Botucatu, Brazil
Email: silke@fmb.unesp.br
‡Ostbayerische Tech. Hochschule, Fakultät Informatik/Mathematik, Regensburg, Germany
Email: christian.hook@hs-regensburg.de
§UNESP - São Paulo State University, School of Sciences, Bauru, Brazil
Email: papa@fc.unesp.br

***Abstract*—Approximately $50,000$ to $60,000$ new cases of Parkinson's disease (PD) are diagnosed yearly. Despite being non-lethal, PD shortens life expectancy of the ones affected with such disease. As such, researchers from different fields of study have put great effort in order to develop methods aiming the identification of PD in its early stages. This work uses handwriting dynamics data acquired by a series of tasks and proposes the application of a deep-driven graph-based clustering algorithm known as Optimum-Path Forest to learn a dictionary-like representation of each individual in order to automatic identify Parkinson's disease. Experimental results have shown promising results, with results comparable to some state-of-the-art approaches in the literature.**

***Index Terms*—Parkinson's disease, Optimum-Path Forest, Handwriting Dynamics**

I. Introduction

The cure for neurodegenerative diseases has been constantly researched by Medicine, mainly with respect to Parkinson's disease (PD), which affects nearly 1 million people only in the United States, and around 7 to 10 million people might be living with PD worldwide. Also, the number of new cases diagnosed each year ranges between $50,000$ to $60,000$ individuals according to the National Parkinson's Foundation [1]. Parkinson's disease is characterized by motor dysfunctions, it is a chronic, progressive and multilesion disease caused by the loss of a neurotransmitter called *Dopamine* [2]. Such illness is usually diagnosed through a clinical exam by a neurologist with expertise in movement analysis. The PD is considered non-lethal, but people with PD have a shorter life expectancy than the general population.

More often in the elderly population, PD produces alterations in gait and posture that may increase the risk of falls and lead to mobility disabilities. As such, it impacts daily activities and reduces the quality of life concerning patients and their families [3]–[5], especially because it does not have cure to date. Drugs known as dopaminergic medications and therapy are currently used to treat PD symptoms, being the *Levodopa* (L-dopa) the most widely used for such purpose. Another treatment that has been widely employed is the Deep Brain Stimulation, which is a surgical procedure that delivers electrical pulses to brain cells in order to reduce the effects of the symptoms.

The science does not measure efforts in order to make the quality of life of PD patients better. In computer science, for instance, techniques such as image processing, neural networks and others have been widely applied in the pursuit of better results in both treatment and diagnosis. Spadotto et al. [6], for instance, introduced the Optimum-Path Forest (OPF) [7], [8] classifier to aid the automatic identification of Parkinson's disease. Later on, the same group proposed an evolutionary-based approach to select the most discriminative set of features that helped improving PD recognition rates [9].

Most works that address automatic PD recognition deal with voice-based data. Procedures to identify voiced and unvoiced (silent) periods have been actively pursued to analyze continuous speech samples, since most techniques that quantify periodicity and regularity in voice signals are applied in the voiced regions only [10]. Das [11] presented a comparison of multiple classification methods for the diagnosis of PD, such as neural networks, regression and decision trees. Several evaluation methods were employed to calculate the performance of the classifiers, being the experiments conducted in a dataset composed of biomedical voice measurements from 31 people, in which 23 were diagnosed with Parkinson's disease. The best results were achieved by neural networks (around 92.9% of PD recognition rate).

Recently, Pereira et al. [12], [13] proposed to extract features from handwritten exams using visual features, which are learned from some drawings the patients were asked to perform, being the data used in the work made available in a dataset called "HandPD"[1]. Later on, Pereira et al. [14] drove its approach to a deep learning application using the signals (time series) captured by the biometric pen *BiSP*® [15], which were further converted to the image domain with different

*Both authors contributed equally.

[1]http://wwwp.fc.unesp.br/~papa/pub/datasets/Handpd/

resolutions and used as input to a Convolutional Neural Network.

Another interesting methodology to learn discriminative features from data is related to the well-known Bag-of-words (BoW), though being quite difficult to establish the size of the bag (dictionary), as well as another open problem is how to choose the words that will compose that bag. Some years ago, Afonso et al. [16] proposed to use the unsupervised OPF [17] to learn proper dictionaries since it does not require the number of words beforehand, thus becoming an useful tool for BoW purposes. Later on, Afonso et al. [18] presented a deep-hierarchical OPF (dOPF) clustering algorithm to make it way more efficient, and validated it in the context of seismic-geological data classification.

Although BoW usage is not new in the context of time series for biomedical purposes [19], to best of our knowledge, it has not been applied for the identification of Parkinson's disease along with graph-based clustering algorithms so far, which turns out to be the main contribution of this work. Another main contribution is to use dOPF to learn dictionaries in a hierarchical way, where different layers of knowledge are used to compose the final dictionary. In short, the main idea of this work is to employ dOPF in the context of BoW applied for Parkinson's disease detection using the time series data from the HandPD dataset. The remainder of this work is organized as follows: Section II describes the theoretical background related to both OPF and dOPF. Our proposed approach is detailed in Section III. The experimental setup, dataset and results are presented in Section IV. Finally, Section V states conclusions and future works.

II. Optimum-Path Forest Clustering

The main problem in unsupervised learning is to identify clusters in a dataset $\mathcal{Z}$, in which samples belonging to the same group should share some level of similarity. The Optimum-Path Forest clustering algorithm handles this problem as a graph partition task, where a competitive process among prototype samples (a subset from $\mathcal{Z}$) offers optimum-cost paths to the remaining samples in order to "conquer" them. The outcome of this competition process is a collection of trees (forest) rooted at each prototype, in which each tree represents a different cluster.

Given the dataset $\mathcal{Z}$, we can create a graph $(\mathcal{Z}, \mathcal{A}_k)$, where $\mathcal{A}_k$ is a k-nearest neighbors adjacency relation, and each sample $\mathbf{x} \in \mathcal{Z}$ encodes a graph node in $\Re^n$, i.e., it basically stands for a feature vector extracted from a dataset sample. Let $d(\mathbf{s}, \mathbf{t})$ be the distance between graph nodes $\mathbf{s}$ and $\mathbf{t}$, being the edge connecting such samples (i.e., $(\mathbf{s}, \mathbf{t})$) weighted by that distance. Also, a given node $\mathbf{s}$ is weighted by a probability density function (pdf) $\rho(\mathbf{s})$ defined as follows:

$$\rho(\mathbf{s}) = \frac{1}{\sqrt{2\pi\sigma^2}|\mathcal{A}(\mathbf{s})|} \sum_{\forall t \in \mathcal{A}(\mathbf{s})} \exp\left(\frac{-d^2(\mathbf{s}, \mathbf{t})}{2\sigma^2}\right), \quad (1)$$

in which $\sigma = \frac{d_f}{3}$, and d_f is the length of the longest edge in $(Z, \mathcal{A}_k)$. The choice of this parameter considers all nodes for density computation since a Gaussian function covers most samples within $d(\mathbf{s}, \mathbf{t}) \in [0, 3\sigma]$. The Parzen-window method is the most common method to estimate a probability density function, and is provided by Equation 1 based on the isotropic Gaussian kernel when the arcs are defined by $(\mathbf{s}, \mathbf{t}) \in \mathcal{A}_k$ if $d(\mathbf{s}, \mathbf{t}) \leq d_f$.

This approach, however, presents some issues with the differences in scale and sample concentration, which can be solved by adaptive choices of d_f depending on the region of the feature space [20]. By choosing the best value for the k-nearest neighbors within [1, k_{max}], for $1 \leq k_{max} \leq |\mathcal{Z}|$, it is possible to tackle both issues of different concentration and scale reduction. Rocha et al. [17] proposed a solution by considering the minimum graph cut provided by clustering results for $k \in [1, k_{max}]$, according to a measurement suggested by Shi and Malik based on graph cuts [21].

Let $\pi_{\mathbf{t}}$ a path in $(Z, \mathcal{A}_k)$ that can be defined as a sequence of adjacent nodes that starts in a root $R(\mathbf{t})$ and ends at a sample $\mathbf{t}$, being $\pi_{\mathbf{t}} = \langle \mathbf{t} \rangle$ a trivial path, and $\pi_{\mathbf{s}} \cdot \langle \mathbf{s}, \mathbf{t} \rangle$ the concatenation of $\pi_{\mathbf{s}}$ and arc $(\mathbf{s}, \mathbf{t})$. The main point is to find a path whose lowest density value is maximum among all possible paths $\pi_{\mathbf{t}}$ with roots on the maxima of the pdf. Thus, each maximum should define an influence zone (cluster) by selecting samples that are more strongly connected to it than to any other maximum. Formally, this process can be defined as the maximization of $f(\pi_{\mathbf{t}})$ for all $\mathbf{t} \in \mathcal{Z}$, such that

$$\begin{aligned} f(\langle \mathbf{t} \rangle) &= \begin{cases} \rho(\mathbf{t}) & \text{if } \mathbf{t} \in \mathcal{S} \\ \rho(\mathbf{t}) - \delta & \text{otherwise} \end{cases} \\ f(\langle \pi_{\mathbf{s}} \cdot \langle \mathbf{s}, \mathbf{t} \rangle \rangle) &= \min\{f(\pi_{\mathbf{s}}), \rho(\mathbf{t})\} \end{aligned} \quad (2)$$

for $\delta = \min_{\forall (\mathbf{s},\mathbf{t}) \in \mathcal{A}_k | \rho(\mathbf{t}) \neq \rho(\mathbf{s})} |\rho(\mathbf{t}) - \rho(\mathbf{s})|$ and $\mathcal{S}$ being a root set with one element for each maximum of the pdf. High values of delta reduce the number of maxima. This work sets $\delta = 1.0$ and scales real numbers $\rho(t) \in [1, 1000]$. In summary, the OPF algorithm maximizes $f(\pi_{\mathbf{t}})$ such that the optimum paths form an optimum-path forest — a predecessor map P with no cycles that assigns to each sample $\mathbf{t} \notin \mathcal{S}$ its predecessor $P(\mathbf{t})$ in the optimum path from $\mathcal{S}$ or a marker nil when $\mathbf{t} \in \mathcal{S}$.

A. Deep-Hierarchical Optimum-Path Forest

One of the main advantages of OPF concerns its capability in computing the number of clusters on-the-fly, which means such information is not required beforehand. On the other hand, its bottleneck is to set the exact number of clusters when one knows that information. One possible solution is to consider different values for the parameter k_{max} in order to reach the desired number of clusters. However, the larger the dataset, the more costly this process will become.

In order to overcome the aforementioned issue, Afonso et al. [18] proposed a multi-level clustering algorithm based on the OPF approach. Each level (layer) computes an OPF through the very same process as previously described using the roots (prototypes) from the OPF computed in the predecessor layer as new inputs. The number of layers is user-defined

and set according to the number of clusters that is supposed to be reached (or close to) in the last layer. Since the prototypes are located in the highest density regions, they are very suitable to represent nearby samples, as argued in the works conducted by Castelo and Calderón-Ruiz [22] and Afonso et al. [23].

Let $\mathcal{S}_i$ be the set of prototypes at layer L_i, $i = 1, 2, \ldots, l$, in which l stands for the number of layers. Since each root will be the maximum of a pdf (Equation 1), we have a set of samples that fall in the same optimum-path tree and are represented by the very same prototype (root of that tree) in the next layer. In summary, the higher the number of layers, the less prototypes (clusters) one shall have, i.e., $|\mathcal{S}_1| < |\mathcal{S}_2| < \ldots < |\mathcal{S}_l| < \ldots \leq 1$. Therefore, at layer l, one shall find only one cluster when $l \to \infty$. Figure 1 displays the OPF-based architecture for deep-driven feature space representation.

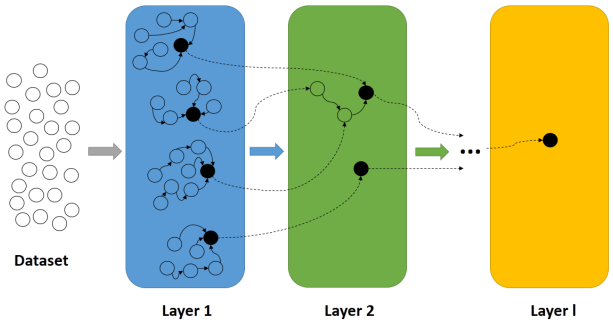

Fig. 1: Architecture of an l^{th}-layered dOPF.

At layer L_1, it is observed four clusters (optimum-path trees), in which the black filled nodes stand for the set of prototypes at that layer, i.e., $\mathcal{S}_1$. Some of these prototypes will become new prototypes at L_2, and others not (we can observe both filled and unfilled nodes at layer 2). This process is carried out up to the l^{th} layer specified by the user. Notice at the coarsest scale, i.e., L_l, we shall find only one cluster. Therefore, the user can halt the process as soon as the number of desired clusters (or close to it) is met.

III. Proposed Approach

This section describes all steps performed in the work to evaluate dOPF and BoW in the context of Parkinson's disease identification, as depicted in Figure 2.

a) Data acquisition: Individuals were submitted to a series of tasks, in which they were asked to perform some hand movements and drawings using a biometric pen that contains six sensors in charge of recording hand movements (Figure 3) [15]. The movements are represented by six different channels: microphone, finger grip, axial pressure of ink refill, and tilt and acceleration in the x, y and z directions.

Figure 4 depicts an example of an exam containing six tasks that evaluate the hand movements and help to detect any anomalies. In the first task (exam (a) in Figure 4), the individual is asked to draw a circle 12 times in the same place without stopping the movement between each circle. In the second task (exam (b) in Figure 4), the individual performs the same movement as in exam (a), but with its hand in the air. The third (exam (c) in Figure 4) and fourth (exam (d) in Figure 4) tasks concern drawing the spirals and meanders, respectively, over a guideline only once from the inner to the outer part. The last two tasks, i.e., exam (e) and exam (f) in Figure 4, stand for the diadochokinese test, which is basically composed of hand-wrist movements performed with both hands. Each exam results in six different datasets, one for each task, and each sample from the dataset corresponds to an array of responses captured by each sensor in the interval of 1 ms.

b) Local descriptor extraction: Given the recorded signals, the local descriptors are computed through a sliding window that goes along each of the six signals and computes a single-level Discrete Wavelet Transform (DWT) in each segment. In fact, since there are six different signals, we work with six sliding windows, in which the segments of time within each of them have always the same initial and final times as they shift along the signals. The size of the sliding window and shifting are both user-defined. The DWT is applied to each segment of time separately, and the results in each segment are concatenated in order to form the final local descriptor[2].

c) Dictionary formulation: The dictionary formulation aims to find the most representative "words" (descriptors) among a set of descriptors from the "bag" that are used in a later step for computing of a new sample representation. This step is usually performed by a clustering algorithm, in which the number of clusters defines the size of the dictionary, and each centroid becomes a "word" of the dictionary. It is usual to play with the size of the dictionary in order to find some trade-off between the computational cost and accuracy rate.

d) The new representation: A signal can be represented by a set of descriptors, which can range from dozens to thousands. Some of these descriptors may be similar or only represent noisy information. Thus, in order to obtain a compressed and meaningful representation of the signal, the descriptors were quantized based on the dictionary computed previously. The quantization step will provide a histogram for each sample with length equals to the size of the dictionary, in which each bin will have the frequency of its closest word in the input signal. Then, the final histogram is further used as an input for machine learning algorithms.

IV. Experiments and Results

The experimental setup used all data recorded from a total of 66 exams, being 35 control individuals and 31 patients. The output of the protocol discussed in the previous section results in six different datasets, one for each task. The dictionary learning step was performed by means of three different techniques: dOPF, k-means[3] and OPF[4]. The main idea is to

[2]We used sliding windows of size 100 ms with stride of 50 ms, being such values empirically chosen.

[3]Our own implementation.

[4]https://github.com/LibOPF/LibOPF

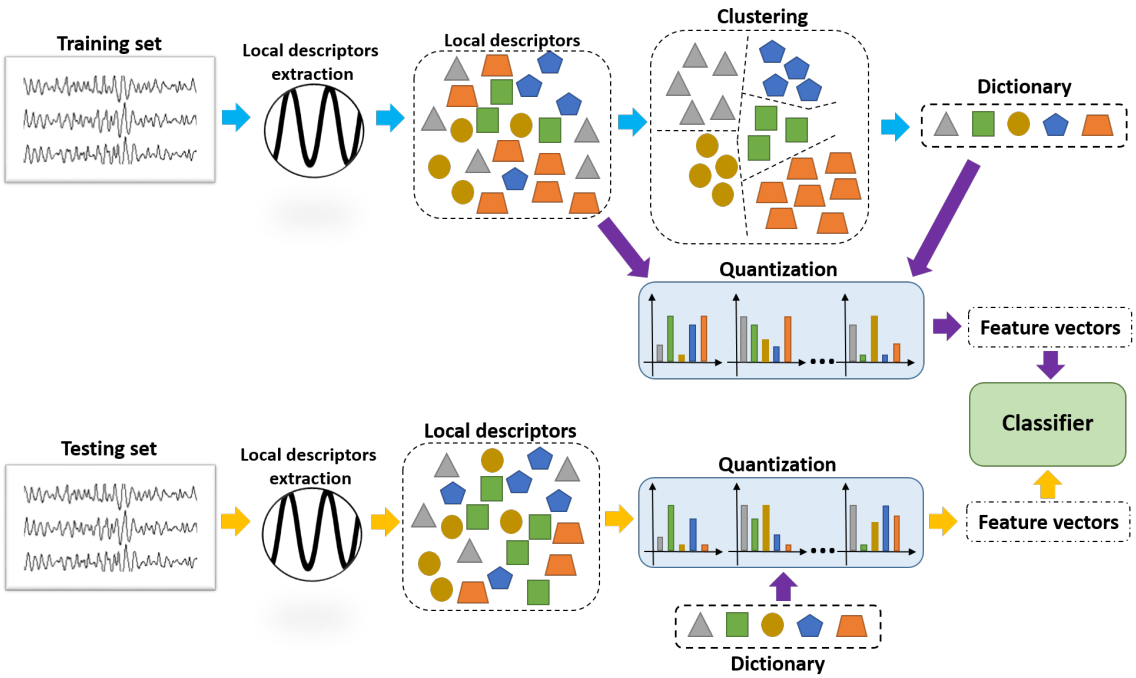

Fig. 2: Proposed approach based on BoW and dOPF for computer-aided PD diagnosis. The main workflow is indicated by the light blue arrows: local descriptors are extracted and clustered in order to build the dictionary. The dictionary is used for the quantization of both training and testing signals that is the process of computing the feature vectors (flow indicated by purple arrows). Similarly to the training phase, testing signals have their descriptors computed and the signals are quantized (flow indicated by yellow arrows). Finally, a classifier is fed by the resulting training and testing feature vectors. Notice the two depicted dictionaries are the same.

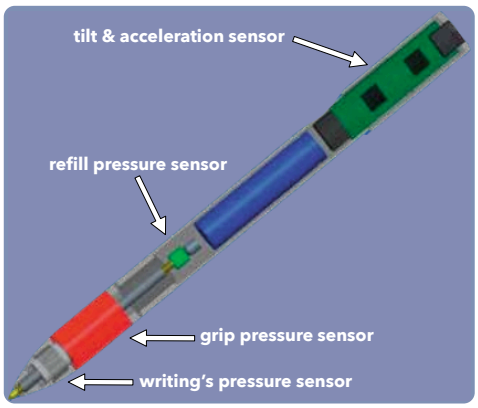

Fig. 3: Biometric pen. Extracted from [14].

evaluate the quality of clustering of each technique through the accuracy rate obtained in the classification phase. The architecture used by dOPF is composed of four layers, in which the values of k_{max} are: 100 for the first layer, 1% of the number of clusters computed in the previous layer are used as an input for the second layer, and 10% of the number of clusters computed in their respectively antecessor layers for the third and fourth layers. The value of k for k-means is always set as the number of clusters found by the fourth (last) layer of dOPF approach. Regarding the OPF algorithm, the values for k_{max} were empirically set as $2,500$ for the Spiral and Meander datasets, and as $1,500$ for the remaining datasets. The idea in using the same number of clusters for dOPF and k-means is to allow a fair comparison between them.

Table I presents the number of descriptors extracted from the training set of each dataset, as well as the number of words computed in each case. In the column regarding dOPF, it is shown the number of words found for each of the four layers, but only the ones computed in the last layer (bolded) are used for the quantization of both training and testing sets.

The experiments were performed using the hold-out procedure with 15 runs. Both training and testing sets were partitioned using 50% of the entire dataset each, being randomly generated in each new run. In this step, there were employed three different classifiers for comparison purposes: Bayesian Classifier (BC)[5], supervised OPF (sOPF)[6] and SVM using a Radial Basis Function (RBF) kernel with parameter optimization (SVM-RBF) [24].

Tables IIa— IIf present the mean recognition rates concerning all six exams, being the accuracy computed according to Papa et al. [7]. The best results are defined according to the

[5]Our own implementation.
[6]https://github.com/LibOPF/LibOPF

TABLE I: Number of descriptors extracted from the training set and number of words computed by each technique.

dataset (task)	# descriptors	Deep-OPF	K-means	OPF
Circ-A exam (a)	18,000	5,682 - 2,584 - 228 - **68**	68	693
Circ-B exam (b)	11,898	538 - 376 - 43 - **17**	17	33
Spiral exam (c)	46,637	12,118 - 3,951 - 370 - **92**	92	1,424
Meander exam (d)	41,094	10,865 - 3,937 - 429 - **99**	99	1,591
Dia-A exam (e)	14,608	666 - 480 - 95 - **47**	47	80
Dia-B exam (f)	13,947	657 - 394 - 78 - **27**	27	70

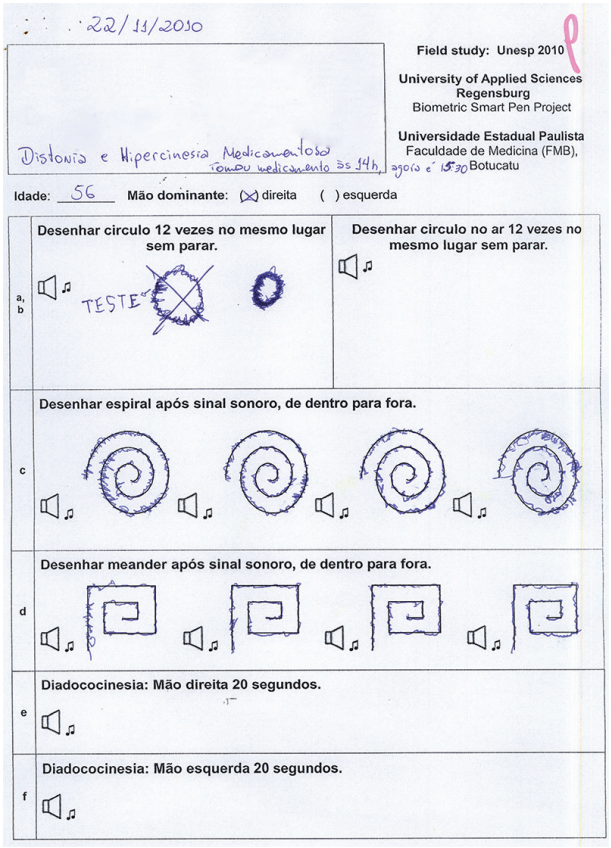

Fig. 4: Form used to assess the handwritten skills. Extracted from [14].

TABLE II: Overall accuracies.

(a) Circ-A dataset.

	BC	sOPF	SVM-RBF
dOPF	**82.96±2.88**	**81.71±5.12**	73.87±4.58
***k*-means**	**83.38±4.22**	**82.01±5.11**	65.80±12.39
OPF	**81.06±4.36**	**81.90±4.89**	76.17±6.92

(b) Circ-B dataset.

	BC	sOPF	SVM-RBF
dOPF	68.75±7.96	69.14±6.95	**77.31±4.45**
***k*-means**	67.80±7.44	65.58±6.79	**74.54±6.39**
OPF	70.81±4.62	**73.08±8.96**	**76.69±5.38**

(c) Spiral dataset.

	BC	sOPF	SVM-RBF
dOPF	**78.30±5.80**	76.73±6.83	77.25±3.46
***k*-means**	73.37±5.37	73.11±5.31	78.83±2.20
OPF	75.40±3.09	75.57±3.13	**81.03±2.40**

(d) Meander dataset.

	BC	sOPF	SVM-RBF
dOPF	73.33±4.97	74.07±2.90	**80.45±2.42**
***k*-means**	76.07±3.31	76.09±2.77	78.26±3.91
OPF	78.53±3.15	77.21±3.52	**81.07±2.60**

(e) Dia-A dataset.

	BC	sOPF	SVM-RBF
dOPF	**69.86±7.21**	**70.93±7.29**	**68.69±7.26**
***k*-means**	**72.18±7.46**	**72.43±5.81**	**73.93±8.66**
OPF	**70.72±6.60**	**67.01±7.45**	**68.69±7.26**

(f) Dia-B dataset.

	BC	sOPF	SVM-RBF
dOPF	**67.96±8.10**	64.86±7.93	61.89±8.49
***k*-means**	**72.92±8.51**	**69.84±9.03**	**67.24±9.31**
OPF	63.77±8.85	67.25±6.80	**66.30±7.38**

Wilcoxon signed-rank [25] with significance of 0.05, which pointed out the best ones in bold for each exam. Further, we also considered the best among all exams as the underlined ones.

Let us first analyze the best results among all. The statistical evaluation pointed out [OPF, SVM-RBF] and [k-means, BC] as the best pairs of [dictionary learner, classifier] with accuracies near to 81% and 83%, respectively. Comparing that recognition rates against some previous works, the proposed approach showed significant gains (from 10% to 30%) against the one presented by Pereira et al. [13]. Despite that our results were slightly below those achieved by a further work of the same authors that makes use of deep learning techniques [14], our approach is way more efficient than using deep learning techniques taking into account a few architectures.

With respect to the best accuracies concerning each exam, dOPF obtained very much suitable results, being more accurate than naïve OPF in most cases. Supervised OPF obtained good results as well, but SVM-RBF achieved the best recognition rates in a few more situations. Additionally, we also evaluated the accuracy per class for all situations, as presented in Tables IIIa— IIIf, whose best results are also highlighted considering the Wilcoxon signed-rank. The best results for each class are in bold, and the best among all datasets is underlined. Actually, the main improvement concerns the accuracy for the identification of healthy individuals, since Pereira et al. [14] obtained recognition rates nearly 50% over the Meander and Spirals datasets for the control class. The proposed approach

increased not only the global accuracy with respect to the work by Pereira et al. [14], but also the specificity and sensitivity for most of the cases. Also, Circ-A dataset provided two out of the five best results, thus showing as a good alternative for the Parkinson's Disease identification.

Table IV presents the mean computational load required by each technique to learn the dictionary. Notice the computational burden for dOPF considers the four layers. In this context, k-means figured as the fastest one due to its simplicity. If one considers dOPF and OPF only, we can observe the former is about 78 times faster in Circ-B dataset, which is quite effective. The lowest gains can be observed in both Meander and Spiral datasets. The small differences come from the fact the value used for k_{max} in both situations is small, thus justifying the fact the dictionaries computed in these datasets have very high dimension when compared to others.

V. Conclusions

This work introduced a deep-hierarchical version of the unsupervised OPF algorithm for dictionary learning in the context of computer-aided Parkinson's disease identification. The experiments were performed using data from handwriting dynamics, similarly to the work by Pereira et al. [14], but now handled as signals and not images.

The application of the BoW paradigm can extract more information by computing local descriptors that can enhance the overall accuracy. Also, dOPF showed satisfactory results in its first application for BoW-based Parkinson's Disease identification. Experiments over six datasets considered dOPF against the well-known k-means and naïve OPF clustering for dictionary learning. Further, supervised techniques were used for classification purposes.

Future works will consider learning hierarchical BoWs, i.e., one bag for each layer in the dOPF formulation. We believe each layer can carry different information about the problem.

Acknowledgments

The authors are grateful to FAPESP grants #2014/12236-1 and #2016/19403-6, Capes, and CNPq grant #306166/2014-3.

References

[1] N. P. Fundation, "The parkinson's foundations moving day walks fund local parkinsons programs in 2017."

[2] A. J. Lees, J. Hardy, and T. Revesz, "Parkinson's disease," *The Lancet*, vol. 373, no. 9680, pp. 2055–2066, 2009.

[3] B. E. Maki and W. E. McIlroy, "Change-in-support balance reactions in older persons: An emerging research area of clinical importance," *Neurologic Clinics*, vol. 23, no. 3, pp. 751–783, 2005.

[4] G. F. Marchetti and S. L. Whitney, "Older adults and balance dysfunction," *Neurologic Clinics*, vol. 23, no. 3, pp. 785–805, aug 2005.

[5] Y. J. Zhao, L. C. S. Tan, P. N. Lau, W. L. Au, S. C. Li, and N. Luo, "Factors affecting health-related quality of life amongst asian patients with parkinson's disease," *European Journal of Neurology*, vol. 15, no. 7, pp. 737–742, jul 2008.

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

[7] 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.

[8] 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.

[9] A. A. Spadotto, R. C. Guido, F. L. Carnevali, A. F. Pagnin, A. X. Falcão, and J. P. Papa, "Improving parkinson's disease identification through evolutionary-based feature selection," in *IEEE International Conference of the Engineering in Medicine and Biology Society*, 2011, pp. 7857–7860.

[10] 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. 07, no. 04, pp. 147–156, 2014.

[11] R. Das, "A comparison of multiple classification methods for diagnosis of parkinson disease," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1568 – 1572, 2010.

[12] C. R. Pereira, D. R. Pereira, F. A. da Silva, C. Hook, S. A. T. Weber, L. A. M. Pereira, and J. P. Papa, "A step towards the automated diagnosis of parkinson's disease: Analyzing handwriting movements," in *IEEE 28th International Symposium on Computer-Based Medical Systems*, 2015, pp. 171–176.

[13] C. R. Pereira, D. R. Pereira, F. A. Silva, J. P. Masieiro, S. A. T. Weber, C. Hook, and J. P. Papa, "A new computer vision-based approach to aid the diagnosis of parkinson's disease," *Computer Methods and Programs in Biomedicine*, vol. 136, pp. 79–88, 2016.

[14] C. R. Pereira, S. A. T. Weber, C. Hook, G. H. Rosa, and J. P. Papa, "Deep learning-aided parkinson's disease diagnosis from handwritten dynamics," in *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2016, pp. 340–346.

[15] G. University of Applied Sciences Team, Regensburg. (2002) A novel multisensoric system recording and analyzing human biometric features for biometric and biomedical applications. [Online]. Available: http://www.bisp-regensburg.de/references.html

[16] L. C. S. Afonso, J. P. Papa, L. P. Papa, A. N. Marana, and A. R. Rocha, "Automatic visual dictionary generation through optimum-path forest clustering," in *2012 19th IEEE International Conference on Image Processing*, 2012, pp. 1897–1900.

[17] L. M. Rocha, F. A. M. Cappabianco, and A. X. Falcão, "Data clustering as an optimum-path forest problem with applications in image analysis," *International Journal of Imaging Systems and Technology*, vol. 19, no. 2, pp. 50–68, 2009.

[18] L. Afonso, A. Vidal, M. Kuroda, A. X. F. ao, and J. P. Papa, "Learning to classify seismic images with deep optimum-path forest," in *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2016, pp. 401–407.

[19] J. Wang, P. Liu, M. F. She, S. Nahavandi, and A. Kouzani, "Bag-of-words representation for biomedical time series classification," *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 634 – 644, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S174680941300089X

[20] D. Comaniciu, "An algorithm for data-driven bandwidth selection," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 281–288, 2003.

[21] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, Aug 2000.

[22] C. Castelo-Fernández and G. Calderón-Ruiz, "Automatic video summarization using the optimum-path forest unsupervised classifier," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, ser. Lecture Notes in Computer Science, A. Pardo and J. Kittler, Eds. Springer International Publishing, 2015, vol. 9423, pp. 760–767.

[23] L. C. Afonso, J. P. Papa, L. P. Papa, A. N. Marana, and A. R. Rocha, "Automatic visual dictionary generation through optimum-path forest clustering," in *19th IEEE International Conference on Image Processing*, 2012, pp. 1897–1900.

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[25] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

TABLE III: Average accuracy rate for each class.

(a) Circ-A dataset.

	BC			sOPF			SVM-RBF		
	dOPF	**k-means**	**OPF**	**dOPF**	**k-means**	**OPF**	**dOPF**	**k-means**	**OPF**
Patient	**83.33±5.62**	**84.17±7.79**	**79.17±12.19**	**77.5±10.24**	**79.58±10.15**	**80.83±8.34**	61.67±15.10	70.42±12.60	75.42±6.87
Control	**82.59±8.09**	**82.59±8.09**	**82.96±8.52**	**85.93±6.59**	**84.44±12.09**	**82.96±6.79**	**67.41±11.67**	**67.04±7.99**	**71.48±9.82**

(b) Circ-B dataset.

	BC			sOPF			SVM-RBF		
	dOPF	**k-means**	**OPF**	**dOPF**	**k-means**	**OPF**	**dOPF**	**k-means**	**OPF**
Patient	**60.83±18.32**	**61.25±11.68**	**74.58±13.39**	**63.75±13.35**	**63.75±13.35**	**54.17±13.04**	**64.58±13.24**	**68.75±14.43**	**57.92±10.07**
Control	76.67±49.12	77.04±12.48	59.99±9.56	71.85±13.98	67.41±12.32	57.04±17.15	**77.04±10.51**	**77.41±6.24**	45.93±7.98

(c) Spiral dataset.

	BC			sOPF			SVM-RBF		
	dOPF	**k-means**	**OPF**	**dOPF**	**k-means**	**OPF**	**dOPF**	**k-means**	**OPF**
Patient	**74.51±10.59**	65.49±10.63	**72.90±6.90**	**78.04±12.07**	67.06±10.14	**71.61±6.27**	67.81±2.17	**73.59±3.67**	**74.58±0.82**
Control	82.08±8.14	81.25±11.33	77.90±6.95	75.42±11.92	79.17±10.48	79.52±6.20	**89.43±1.83**	84.85±2.25	86.43±1.09

(d) Meander dataset.

	BC			sOPF			SVM-RBF		
	dOPF	**k-means**	**OPF**	**dOPF**	**k-means**	**OPF**	**dOPF**	**k-means**	**OPF**
Patient	**76.61±4.04**	69.46±5.96	**76.77±8.44**	**75.38±4.62**	73.23±4.51	**77.85±3.93**	74.81±2.18	71.06±2.62	74.54±1.37
Control	73.33±4.97	82.67±4.88	80.29±4.68	72.76±5.47	78.95±4.99	76.57±5.47	**85.80±0.89**	**84.43±3.76**	87.99±0.72

(e) Dia-A dataset.

	BC			sOPF			SVM-RBF		
	dOPF	**k-means**	**OPF**	**dOPF**	**k-means**	**OPF**	**dOPF**	**k-means**	**OPF**
Patient	**67.50±16.01**	**66.67±14.01**	**51.67±17.15**	**65.83±13.28**	**70.42±12.59**	**52.08±11.10**	**66.25±13.46**	**66.25±15.61**	**47.50±12.87**
Control	**72.22±9.51**	**75.19±5.69**	**50.74±9.70**	**78.52±8.33**	**74.44±6.97**	**55.56±12.67**	**75.19±7.55**	**67.78±12.37**	**50.00±13.46**

(f) Dia-B dataset.

	BC			sOPF			SVM-RBF		
	dOPF	**k-means**	**OPF**	**dOPF**	**k-means**	**OPF**	**dOPF**	**k-means**	**OPF**
Patient	**63.33±12.88**	60.83±11.29	60.00±32.18	**72.50±10.89**	**67.08±13.39**	47.92±10.62	**71.25±11.81**	**73.75±11.23**	**52.08±15.92**
Control	**72.59±9.62**	**68.89±9.89**	50.37±28.07	**73.33±9.56**	**72.59±10.63**	48.52±12.42	**56.29±15.56**	60.74±11.38	**53.70±9.44**

TABLE IV: Dictionary learning computational load [s] required by each technique.

dataset (task)	**dOPF**	**k-means**	**OPF**
Circ-A (a)	968.167	37.008	49,087.137
Circ-B (b)	419.498	13.113	32,777.539
Spiral (c)	6,063.205	239.859	6,643.906
Meander (d)	5,003.233	208.443	5,168.819
Dia-A (e)	613.109	19.878	41,189.133
Dia-B (f)	569.053	11.025	39,367.844