Daubechies Wavelet Cepstral Coefficients for Parkinson's Disease Detection

Soumaya Zayrit
Taoufiq Belhoussine Drissi
Abdelkrim Ammoumou
Laboratory Industrial Engineering, Information Processing and Logistics
University Hassan II, Faculty of Science Ain Chok, Casablanca, Morocco

Benayad Nsiri
Research Center STIS, M2CS Mohammed V University in Rabat
Higher School of Technical Education of Rabat (ENSET), Morocco

The aim of this paper is to evaluate the performance of the approach that focuses on support vector machine (SVM) classification of vocal recording to differentiate between patients affected by Parkinson's disease (PD) and healthy patients. Our study was based on the condition of 38 patients, some of whom are healthy and others who suffer from PD. The study was carried out as follows: The extraction of cepstral coefficients was reached through the transformation of the speech signal by discrete wavelet transform (DWT) and also through cepstral analysis by using the mel scale. At the end, a classification was done by the use of the two kernels linear and radial basis function (RBF) of the SVM classifier.

Keywords: Parkinson's disease; Daubechies wavelet; MFCC; SVM

1. Introduction

The nervous system controls all the workings of the body. The nerves, spinal cord and brain are the components of this system, but if one of them is affected, then the person could have trouble moving, swallowing, breathing, learning or speaking. More than that, it may cause problems with memory, senses or mood.

Among the neurological diseases are degenerative diseases, where nerve cells are damaged or dead, such as the case of Alzheimer's disease and Parkinson's disease.

In the case of Parkinson's disease, the dopamine-producing neurons that are damaged are in the area of the brain called *substantia nigra*. Dopamine is a chemical messenger or a neurotransmitter that plays a role in sending messages to the part of the brain that controls movement and coordination. The loss of this neurotransmitter makes it harder for people to control their movements.

https://doi.org/10.25088/ComplexSystems.29.3.729

Signal treatment has been a means used to determine the diagnosis of diseases. Image treatment is used in Alzheimer's diagnosis [1, 2]. Deepak Gupta and Arnav Julka work on image treatment with the aim of diagnosing Parkinson's disease. To evaluate the proposed model, Parkinson's speech with multiple types of sound recordings and Parkinson's handwriting samples datasets are used. The proposed algorithm can be used in predicting Parkinson's disease with an accuracy of approximately 94% [3].

In vocal analysis, the voice signal represents the convolution between the excitation source and the vocal tract filter. In the spectral domain, this convolution becomes a product that makes it difficult to separate the contribution of the source and the tract. This problem can be overcome by cepstral analysis using cepstral deconvolution transforms, which transform the product to a sum. Cepstral analysis has been widely used for voice recognition [4], specifically mel frequency cepstral coefficients (MFCCs). MFCC analysis consists of exploiting the properties of the human auditory system by the transformation of the linear scale of frequencies in the mel scale and also in the detection of diseases such as Parkinson's disease [5, 6] by using voice recordings from different people during the pronunciation of sustained vowel /a/, then extracting the MFCCs. Respiratory pathologies using pulmonary acoustic signals, the respiratory sounds used in this study, were obtained from the RALE database in the feature extraction stage. The MFCC features are extracted from the respiratory sound signals, and these are fed to the SVM and k-nn classifiers separately in the classification stage. The maximum classification accuracies for the SVM and k-nn classifiers were found to be 92.19% and 98.26%, respectively [7].

Recent studies have been done by T. Belhoussine Drissi et al. in which they work with the discrete wavelet transform (DWT), mel frequency cepstral coefficients and the SVM classifier [8]. Z. Soumaya et al. apply the Daubechies db2 in the 3 scale from which they extract the MFCC with two kernels of SVM linear and radial basis function (RBF) [9]. S. Zayrit et al. [6] propose a hybrid method based on the time frequency domain properties and the K-nearest neighbor in the PD diagnosis. In this paper the MFCCs are extracted from the speech signals through the sorts of DWTs that were tested, and through the cepstral analysis, and at the end applying the support vector machine (SVM) as classifier.

A diagnosis model of Parkinson's disease will be presented in this paper, focusing on the transformation of the speech signal through the proposal of time frequency treatment, followed by a cepstral analysis in order to extract the mel frequency cepstral coefficients. This model has been applied to a database [10] that is composed of 38 patients: 20 sick and the rest healthy. To conclude our study we devised a classification of the database that was done in the following way: the creation of a training base that represents 73% of the database, then testing of the entire database using two kernels linear and radial basis function (RBF) of the SVM classifier.

2. Wavelet Transform

The short-term Fourier transform (STFT) has been proposed to allow time and frequency analysis of a signal (spectrogram calculation). However, this tool has a limitation related to the rigidity of its temporal and frequency resolutions.

There is, however, a problem called the "uncertainty principle": It is not possible to know exactly which frequency exists for a given moment but only which frequency band exists over a time interval. We cannot know exactly which component spectral exists at a given moment. The best we can do is look for what spectral components exist over a given time interval. It is a problem of resolution, and this is the main reason that researchers went from STFT to wavelet transform (WT). Indeed, the STFT gives a fixed resolution for all moments of time, while the WT gives a changeable resolution. The mathematical equation explaining the continuous wavelet transform (CWT) of the signal $s(t)$ is [11]:

$$Ws\left(a, b\right) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} s(t)\psi^{*}\left(\frac{t-b}{a}\right)dt. \tag{1}$$

Here $\psi(t)$ is the mother wavelet, $\psi^{*}(t)$ is the conjugate complex $\psi(t)$, a is the scale factor, and b is the translation parameter. In this paper we will only use the wavelets of Daubechies, which give the best results [8]. Daubechies WT as described by Ingrid Daubechies is an orthogonal wavelet family characterized by a maximal number of vanishing moments of some given support. Figure 1 shows a plot of a Daubechies wavelet function.

The DWT is the discrete version of the CWT. It is implemented using the Mallat algorithm using multi-resolution analysis. This algorithm is based on the definition of a pair of H and G filters that are shown in Figure 2, also called quadratic mirror filters (QMF), whose impulse responses h and g must satisfy certain conditions.

https://doi.org/10.25088/ComplexSystems.29.3.729

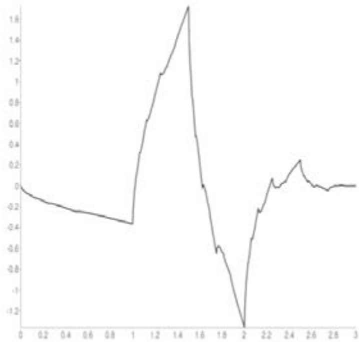

Figure 1. Daubechies wavelet function ψ.

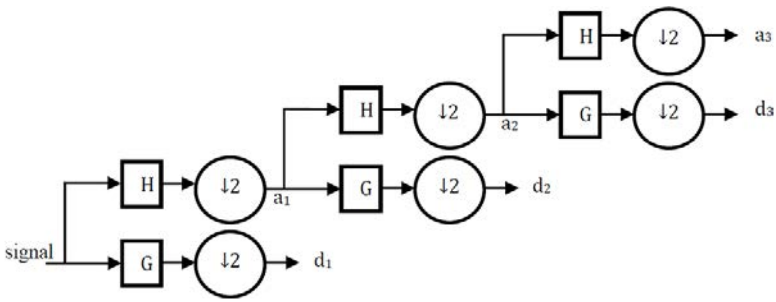

Figure 2. Multi-resolution analysis at three levels of scales (a_i: approximations and d_i: details).

3. Cepstral Coefficients at Mel Scale

Cepstral analysis is widely used in speech and speaker recognition applications, especially the MFCCs. This analysis explores the human auditory system properties, realized through the transformation of the linear scale of frequencies in the mel scale, which is linear in low frequencies and logarithmic in high frequencies [4, 8].

In general, the signal that is the output of a system is caused by the input excitation and also the response of the system. As a consequence, we can consider the signal as a convolution of the input excitation and the response of the system.

Many speech applications necessitate a separate estimate of these individual components, hence a deconvolution between the excitation source and the vocal tract filter is important. Cepstral deconvolution converts the product of two spectra to a sum of two signals.

That is the spectrum of the vocal signal $S = E \cdot V$, with E the spectrum of the source excitation and V the vocal tract. So $\log(S) = \log(E) + \log(V)$, separating V and E. The exploitation of the properties of the human auditory system by a transformation of the frequencies linear scale into mel scales is reached by applying the analysis of mel frequency cepstral coefficients. This last scale is coded through a bank of 15 to 24 triangular filters spaced linearly up to 1 kHz, then logarithmically spaced to the maximum frequencies. The extraction of mel frequency cepstral coefficients is given by (see Figure 3):

$$\mathrm{mel}(f) = 2595 \cdot \log\left(1 + \frac{1}{1 + 700}\right). \tag{2}$$

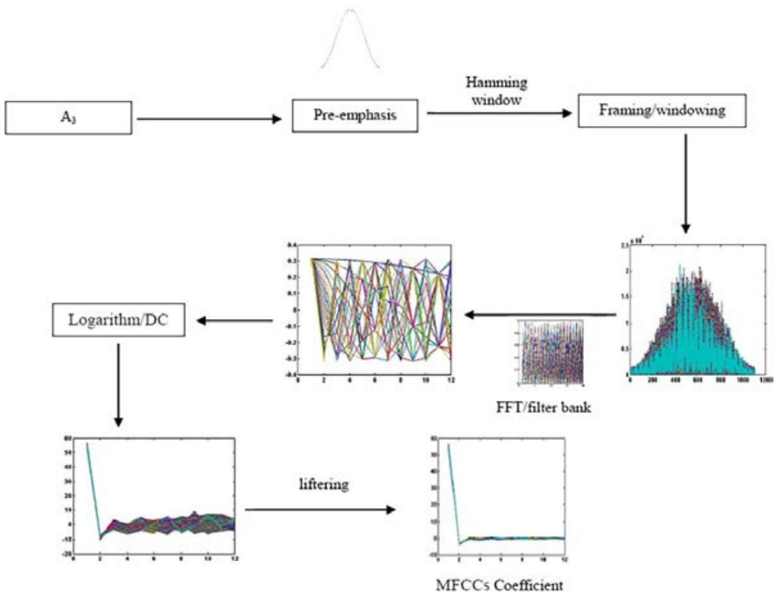

Figure 3. MVCCs extraction.

We calculate the mel-frequency cepstral coefficient from the fast Fourier transform (FFT) coefficients by converting every frame of N samples into the frequency domain instead of the time domain. They are then filtered, employing a triangular bandpass filter bank where $\mathrm{mel}(f)$ is the logarithmic scale of the normal frequency scale f, to obtain MFCCs we convert to time through the discrete cosine transform. The final liftering phase aims to raise the cepstrum, which consequently increases the amplitudes so that they become quite similar.

4. The Support Vector Machine

SVM modeling is based on a discriminative method. The separation between learning samples of two categories by a hyperplane that maximizes the distance between the samples is the principle of the SVM. The linearly separable data and nonlinearly separable data are the two cases of the SVM models. As for the case of the nonlinearly separable data, we change it by projecting the data into a higher dimension so that it will be considered linearly separable. This nonlinear function is reached through a function called kernel function. We represent among those kernels:

- Linear kernel (simple scalar product): $K(x, x_i) = x * x_i$.
- Radial basis function (RBF) kernel: $K(x, x_i) = \exp(-\gamma(x - x_i)^2)$.
- Polynomial kernel: $K(x, x_i) = (x * x_i + C)^2$.

The principal of the SVM classifier is quite simple. It is similar to teaching a child to differentiate between two objects, such as types of fruit, by showing them pictures of apples and oranges during the learning phase while telling them their names. Then we show the child a picture of one fruit and ask them to name it. The aim of this test phase is to make sure that the child is able to say correctly which is which. The algorithm of the SVM classifier follows:

1. Give the learning samples of the two categories $S(x_1, x_2, \ldots, x_N)$, having the label $(y_1, y_2, \ldots, y_N)$ with $(y_i) = 1$ or 0.
2. Input: give the kernel in which we will work then the learning data with the label.
3. Output: obtain the training model: two categories separated by the hyperplane.
4. Input: inject the test sample x_0.
5. Output: obtain the predicted label x_0, which is equal to 1 or 0 depending on the category in which the sample belongs.

5. Methodology and Results

In this paper we will deal with the database [10] with the aim of detecting healthy patients and those affected by PD. First we perform a time-frequency treatment by the Daubechies wavelet in the third scale; this choice is based on the study in [8]. After testing all the DWTs on the acoustic signal, then following the method shown in Figure 4(a), we concluded that the Daubechies wavelet in the third scale and the use of the SVM linear kernel give the best results. In this paper a classification will be effected by the use of two SVM kernels, the linear kernel and the RBF kernel, as given in Figure 4(b). Figure 4 shows the process of signal treatment in both studies.

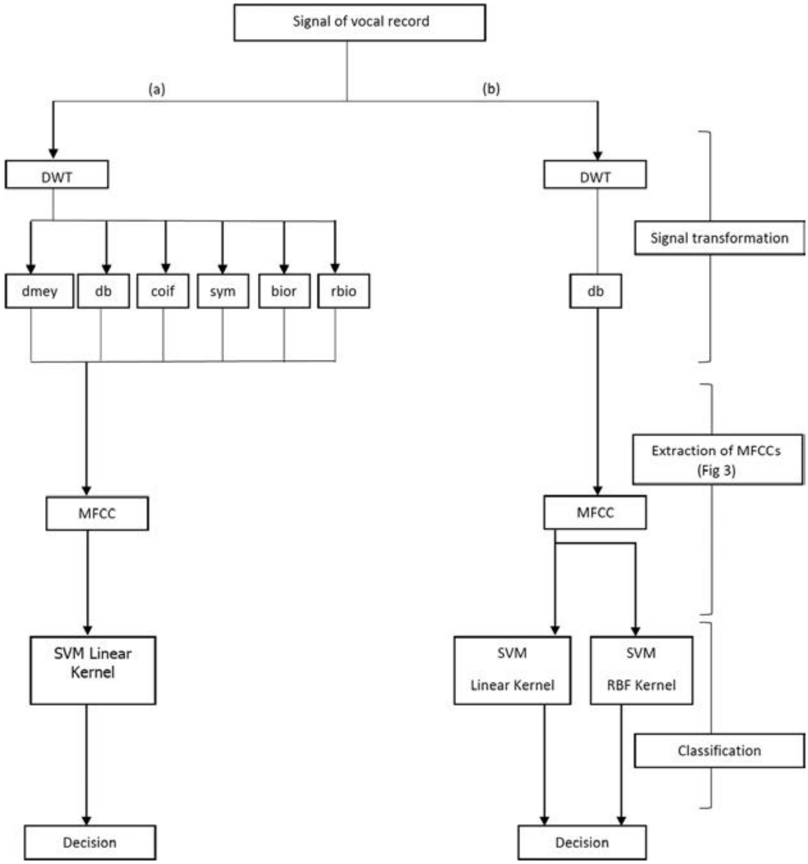

Figure 4. (a) The process followed in [8]; (b) the process followed in this paper.

The algorithm of the DWT stresses the definition of a pair of filters H (lowpass filter) and G (highpass filter). The filter outputs are subsampled by a factor of 2. The highpass filter provides DWT coefficients or signal details at a given scale. The lowpass filter gives the coefficients of the approximation of the signal at the same scale. The recordings in which the patients' voices utter the vowel sound "a" will be transformed into an acoustic signal; after that it is transformed again by the use of Daubechies WT at level two on the scale tree.

Afterward, to obtain the first 12 MFCCs of every patient, we will take the output of the block DWT, in which only the approximation a3 is taken into account. Finally this approximation a3 will be inserted in the MFCC block, in which we use the program "Htk mfcc matlab" [12]. Based on these cepstral coefficients, a categorization will be effected by the use of the SVM classifier, so as to have a precise detection. The MFCC consists of numerous frames that necessitate significant processing time for categorization and that prevent an exact detection [8]. To cope with this problem, the average value of these images is calculated to reach the voiceprint (see Figure 5). To classify healthy and sick patients, we use a training base that accounts for 73% of the database, which contains 20 recordings of sick patients and 18 of healthy patients. Then we do a diagnostic test on the entire data. First, we use the linear kernel of the SVM classifier. Then we do a second diagnostic test while using the RBF kernel. We calculate some measures like accuracy, sensitivity and specificity employing the formula below in order to measure the SVM performance [8]:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{4}$$

$$\text{Specificity} = \frac{TN}{TN + FP}. \tag{5}$$

FN is a false negative: healthy patients who were incorrectly classified, and *TP* is a true positive: healthy patients who were correctly classified. *FP* is a false positive: the PD patients who were incorrectly categorized, and *TN* is a true negative: the PD patients who were correctly categorized. The calculation of the percentage accuracy, sensitivity and specificity of the entire set of recordings from the

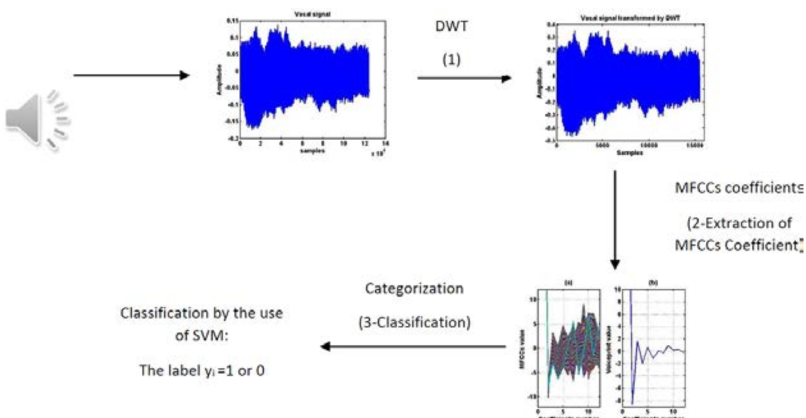

Figure 5. Parkinson's disease diagnosis process.

training base, which was made between the input of the SVM block and the MFCC block output (accounts for 73% of the database) using the linear kernel, and then their percentage by using the RBF kernel, are given in Figure 6 and results without using the DWT are given in Figure 7.

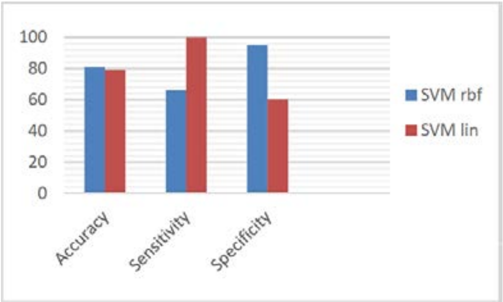

Figure 6. The classification results by the use of DWT and kernel SVM.

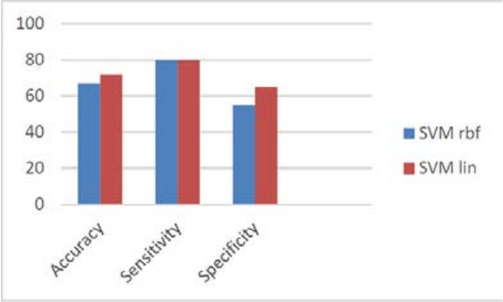

Figure 7. Results of classification without the use of DWT.

6. Conclusion

This paper has come up with a sample of Parkinson's disease detection based on a cepstral analysis after using a signal transformation by the time-frequency treatment by the discrete wavelet transform (DWT) applying this sample on a database of voice recordings of patients while uttering the vowel "a." Daubechies wavelet was used in order to transform the vocal signals by the third-scale approximation. The extraction of the first 12 mel frequency cepstral coefficients (MFCCs) was realized by the insertion of the approximation a3 into the MFCC block. Those cepstral coefficients are employed in the classification applying the support vector machine (SVM) with two kernels linear and radial basis function (RBF). When we do the test with the training base, which accounts for 73% of the database, using the linear kernel, we obtain an accuracy of 79%, whereas the test while using the RBF kernel gives an accuracy of 81%. From that we notice that the RBF kernel is more accurate than the linear kernel, and also by using the wavelet we got better results than the method without the wavelet: 7% higher using the linear SVM kernel and 14% higher using the RBF SVM kernel.

References

[1] R. Cui and M. Liu, "RNN-Based Longitudinal Analysis for Diagnosis of Alzheimer's Disease," *Computerized Medical Imaging and Graphics*, **73**, 2019 pp. 1–10. doi:10.1016/j.compmedimag.2019.01.005.

[2] F. Li and M. Liu, "Alzheimer's Disease Diagnosis Based on Multiple Cluster Dense Convolutional Networks," *Computerized Medical Imaging and Graphics*, **70**, 2018 pp. 101–110. doi:10.1016/j.compmedimag.2018.09.009.

[3] D. Gupta, A. Julka, S. Jain, T. Aggarwal, A. Khanna, N. Arunkumar, V. H. C. de Albuquerque, "Optimized Cuttlefish Algorithm for Diagnosis of Parkinson's Disease," *Cognitive Systems Research*, **52**, 2018 pp. 36–48. doi:10.1016/j.cogsys.2018.06.006.

[4] A. Hacine-Gharbi, "Sélection de paramètres acoustiques pertinents pour la reconnaissance de la parole," Ph.D. thesis, Université d'Orléans, France, 2012.

[5] M. K. MacPherson, J. E. Huber and D. P. Snow, "The Intonation-Syntax Interface in the Speech of Individuals with Parkinson's Disease," *Journal of Speech, Language, and Hearing Research*, **54**(1), 2011 pp. 19–32. doi:10.1044%2F1092-4388(2010%2 F09-0079).

[6] Z. Soumaya, B. D. Taoufiq, N. Benayad, B. Achraf and A. Ammoumou, "A Hybrid Method for the Diagnosis and Classifying Parkinson's Patients Based on Time-Frequency Domain Properties and K-nearest Neighbor," *Journal of Medical Signals & Sensors*, **10**(1), 2020 pp. 60–66. doi:10.4103/jmss.jmss_61_18.

[7] R. Palaniappan, K. Sundaraj and S. Sundaraj, "A Comparative Study of the SVM and k-nn Machine Learning Algorithms for the Diagnosis of Respiratory Pathologies Using Pulmonary Acoustic Signals," *BMC Bioinformatics*, **15**(1), 2014 223. doi:10.1186/1471-2105-15-223.

[8] T. Belhoussine, S. Zayrit, B. Nsiri and A. Ammoummou, "Diagnosis of Parkinson's Disease Based on Wavelet Transform and Mel Frequency Cepstral Coefficients," *International Journal of Advanced Computer Science and Applications*, **10**(3), 2019 pp. 125–132. doi:10.14569/IJACSA.2019.0100315.

[9] Z. Soumaya, B. D. Taoufiq, B. Nsiri and A. Abdelkrim, "Diagnosis of Parkinson Disease Using the Wavelet Transform and MFCC and SVM Classifier," in *Proceedings of 2019 IEEE World Conference on Complex Systems (WCCS'19)*, Ouarzazate, Morocco (M. Essaaidi and M. Nemiche, eds.), Piscataway, NJ: IEEE, 2019 pp. 1–6.

[10] 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*, **17**(4), 2013 pp. 828–834. doi:10.1109/jbhi.2013.2245674.

[11] C. K. Chui, *An Introduction to Wavelets*, Boston: Academic Press, 1992.

[12] K. Wojcicki. "HTK MFCC MATLAB." MATLAB Central File Exchange. (Jul 29, 2020). www.mathworks.com/matlabcentral/fileexchange/32849-htk-mfcc-matlab/content/mfcc/writehtk.m.