مجلة جامعة دمشق للعلوم ....
المجلد ... العدد... : الصفحات.
Damascus university journal
V… ( ): ??-??

تصنيف الاضطرابات الصوتية

د. * تاولي سيدي أحمد
قسم الهندسة البيوطبية، مختبر أبحاث الهندسة البيوطبية
كلية التكنولوجيا، جامعة أبوبكر بلقايد الجزائر
sidiahmed.taouli@univ-tlemcen.dz

الملخص:

تقدم هذه المقالة تصنيف الأصوات المرضية، وهو مجال حيوي في الصحة الصوتية.

يقدم عملنا منهجية تدمج معالجة الإشارات، بما في ذلك تحليل المويجات واستخراج المتغيرات، ويكشف عن نتائج واعدة تم تحقيقها باستخدام نموذج نموذج آلة دعم المتجهات للتصنيف. الهدف الأساسي هو تعزيز تشخيص وإدارة الاضطرابات الصوتية من خلال توفير أدوات أكثر فعالية لأخصائيي الرعاية الصحية، استنادًا إلى قاعدة بيانات(Voiced) .يؤكد هذا العمل على الأهمية الحاسمة للصحة الصوتية وضرورة الاستثمار في طرق تشخيصية أكثر دقة وسهولة في الوصول إليها، مع فتح آفاق جديدة للرعاية وجودة حياة الأفراد المصابين.

الكلمات المفتاحية: التصنيف، الأصوات المرضية، الإشارات الصوتية، تحليل المويجات, استخراج المعلّمات، نموذج SVM، قاعدة بياناتVOICED.

تاريخ الايداع
تاريخ القبول

Damascus university journal
V… ( ): ??-??

مجلة جامعة دمشق للعلوم ….
المجلد … العدد… : الصفحات.

Classification of Voice Disorders

Dr: *Sidi.Ahmed Taouli

Department of biomedical genius, biomedical engineering research laboratory

Faculty of technology, University Aboubekr-Belkaid, algeria

sidiahmed.taouli@univ-tlemcen.dz

Abstract:

This article introduces the classification of pathological voices, a vital area in vocal health. Our work presents a methodology integrating signal processing, including wavelet analysis and parameter extraction, and unveils promising results achieved using a Support Vector Machine (SVM) model for classification. The primary objective is to enhance the diagnosis and management of vocal disorders by providing more effective tools to healthcare professionals, based on the VOICED database. This work underscores the critical importance of vocal health and the necessity of investing in more precise and accessible diagnostic methods, while also opening up new prospects for care and the quality of life of affected individuals.

Keywords: Classification, Pathological Voices, Vocal Signal, Wavelet Analysis, Parameter Extraction, SVM Model, VOICED Database.

.

Received:
Accepted:

2.1Introduction:

Brief Speech processing is a vast area of research that requires the intervention of experts from several specialties. Despite the remarkable development of computer tools and programs, voice-controlled systems have only become successful in recent years. Early detection and precise of vocal pathologies is crucial for appropriate medical intervention. Traditionally, this detection relies on the clinical expertise of health professionals, who subjectively assess vocal alterations. However, this approach has limitations in terms of objectivity, eproducibility and precision. The advent of machine learning technologies provides an unprecedented opportunity to develop more objective and accurate pathological voice classification methods.

Thus, the central question guiding our research is: How can advances in machine learning be used to improve the detection and classification of pathological voices?

Advances in learning and classification technologies have significantly simplified the development of diagnostic methods and tools for vocal conditions.

The automatic detection and classification of pathologies is a current field and still explored by the research community (Malak & Ghulam & Mansour, p. 571) A wide range of acoustic parameters was used for pathology detection, namely pitch, jitter, shimmer,,p harmonic to noise ratio (HNR: Harmonics to Noise Ratio), normalized noise energy (NNE: Normalized Noise). Energy), cepstral coefficients (MFCC: Mel-Frequency Cepstral Coefficients), etc. In the field of automatic voice pathology detection, various classifiers have been proposed such as multi-layer perceptron, Gaussian mixture model, probabilistic neural network, linear discriminant analysis, k-nearest neighbor classifier (KNN: K-Nearest Neighborhood), support vector machines (SVM: Support Vector Machine), etc.

In (Lotfi & Haytham & et Adnène, 2009, p.3) presented a method which is based on the use of a multilayer neural network for the detection of voice pathologies. The results highlight that the pitch (fundamental frequency) and the first three formants prove to be the most effective input parameters for the distinction and identification of voices affected by pathologies, thanks to the use of networks of neurons. (Nesrine & Amina, 2016, p. 26) uses a technique that relies on combining the continuous wavelet transform with higher order statistics. Classification is then carried out using support vector machines (SVM).

(Hammami& Salhi & Labidi, 2020, p. 162) developed a method to identify pathological voices based on higher order features derived from analysis of empirical modal decomposition and discrete wavelet transform. The methodology of this study is divided into three major steps. First, speech signals undergo empirical mode decomposition (EMD), followed by discrete wavelet transform (DWT). From the coefficients from the DWT, several characteristics relevant features are extracted, including higher order features such as skewness, kurtosis and variance, as well as other features such as mean value, energy and entropy. These features are then used to form data vectors representing the speech signals. Finally, an SVM classifier is used for the detection and classification of voices affected by pathologies. The results presented in (MAROUA & RAHMA,2020, p. 25) show the effectiveness of SVM classifier to accomplish pathological voice detection and classification. The accuracy rate obtained using the SVM method and the use of Mel-Frequency Cepstral Coefficients (MFCC) characteristics are considered satisfactory for the database used. On the other hand, (BOUDJELLABA & BOUDJERIDA, 2021, p. 64) presents a method that sets up a system for the identification of vocal disorders using machine learning techniques. They proposed two classification methods, namely Support Vector Machine (SVM) and K Nearest Neighbour (KNN), to evaluate their effectiveness in detecting and classifying various vocal pathologies. This system could be used by speech therapists to perform an objective assessment of their patients' voices, based on acoustic and aerodynamic measurements such as Mel-Frequency Cepstral Coefficients (MFCC), Jitter, Shimmer and Harmonic-to-Noise Ratio (HNR).

As a result of the analysis of the results obtained, it was observed that this method is very effective for the detection of pathological and normal voices, as well as for the classification of different types of pathologies. The central goal of our work is to significantly improve the diagnostic and treatment capabilities of voice disorders. We aim to open new perspectives for healthcare professionals by providing them with more precise and objective tools to assess and manage these conditions. At the same time, we seek to raise awareness of the vital importance of vocal health and highlight the pressing need to invest in more accessible and accurate diagnostic methods. Our commitment focuses specifically on the effectiveness of voice classification pathologies as a promising approach to improve the quality of care and the quality of life of the individuals concerned.

2.2Literature Review:

2.1 Discrete wavelet transform

Describe any materials you used in your research, and methods developed in your research. Discrete wavelet, also known as discrete wavelet transform (DWT), represents an analysis method applied to discrete signals. Compared to the continuous wavelet transform(CWT).

The steps to perform a simple TOD on a discrete signal are:

- Selection of the base wavelet: We choose a base wavelet which will serve as a pattern for the transformation. Wavelets like Haar, Daubechies, and many others are commonly used.
- Decomposition steps:
  - ➢ Step 1: Calculation of approximation and detail coefficients: The first step consists of calculating the approximation coefficients (describing the large-scale components) and the detail coefficients (describing the fine details) from the original signal.
  - ➢ Step 2: Downsampling: The calculated approximation coefficients are then downsampled by removing some samples to achieve lower resolution.
  - ➢ Step 3: Repetition: The previous steps are repeated on the approximation coefficients obtained at each step to obtain approximation and detail coefficients at different scales.
- Repetition and scaling pyramid: The decomposition steps are usually repeated several times to obtain a scaling pyramid of approximation and detail coefficients at different resolutions
- The mathematical formula of DWT involves convolution and subsampling of the parent wavelet and the input signal. Here is how this can be expressed generally:

$$DW(f,\varphi) = \sum_{k} a_{k,j} \cdot \varphi_{j,k}(t) + \sum_{j,k} d_{j,k} \cdot \emptyset_{j,k}(t)$$

Or :

$f(t)$ is the input signal.

- ak,j are the approximation coefficients at scale j and position k.
- dj,k are the detail coefficients at scale j and position k.
- $\varphi j,k(t)$ is a dilated and translated version of the parent wavelet $\varphi(t)$.
- $\emptyset j,k(t)$is a dilated and translated version of the scaling function $\emptyset(t)$, which is related to the mother wavelet and is used to calculate the approximation coefficients.

The coefficients $ak,j\ et\ dj,k$ can be calculated using convolution filters and downsampling operations.

2. 2 Support vector machines

SVMs (Support Vector Machines) were invented by scientists Vladimir Vapnik and Alexey Chervonenkis in the 1960s. They have become a popular tool for discriminative classification. An exciting area of recent application of SVMs is in speech processing. These models have a distinctly different modeling strategy in detecting voice disorders, compared to other classification methods reported in the literature (Godino Llorente, &

Gómez-Vilda &. Sáenz-Lechón & Blanco-Velasco & CruzRoldán, Ferrer-Ballester & Angel, 2005, p. 222). Supervised learning models called SVM are used to separate data points into different groups (classes) by determining an optimal hyperplane in a higher-dimensional feature space. The goal is to discover a dimensional hyperplane that maximizes the distance between data points of various classes. In its elementary form, when the two classes are linearly separable as illustrated in Figure. 1. this approach aims to identify a discriminating hyperplane represented by the equation next:

$$w.x+b \;\; =0$$

The elements of this equation are defined as follows:

- w: Vector of weights ($w1$, $w2$, $w3$, ..., wn).
- x: Vector of attributes ($x1$, $x2$, $x3$, ..., xn).
- b: Threshold of the linear separator.

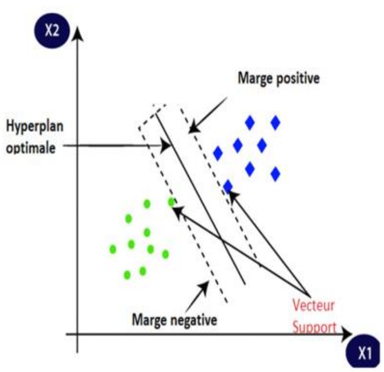

Figure 1- Example of classification by SVM

- ✓ Hyperplane: The hyperplane divides the data into different classes. The task is to find the ideal hyperplane that minimizes classification errors while maximizing the margin between data points of different classes.
- ✓ Margin: The distance between the separation hyperplane and the closest support vectors is represented by the margin of an SVM, and the optimization aims to find the hyperplane that maximizes this distance in order to obtain better separation between classes and better generalization of predictions.
- ✓ Support vector: The data points closest to the separation hyperplane between the different classes are called support vectors. The position and orientation of the optimal hyperplane are determined by these vectors.

2.3 Proposed method

Our method goes through five steps; the following figure illustrates the block diagram of the classification of speech signals.

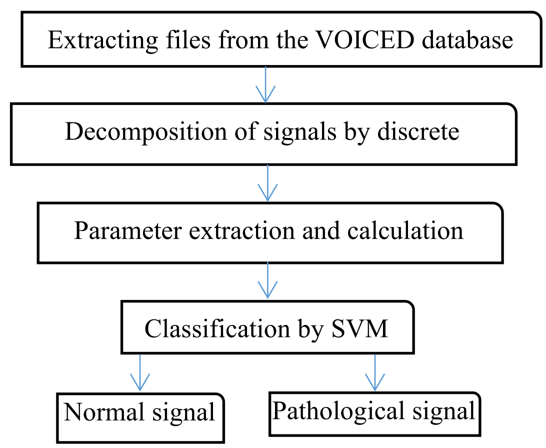

Figure 2 - block diagram of the classification of speech signals

a. Database:

The proposed algorithm is tested on the physionet database.This database includes 208 voice samples, from 150 pathological, and 58 healthy voices (Cesari & De Pietro & Marciano & Niri & Sannino, and Verde, 2018, p. 310). In detail, there were 73 male and 135 female participants. There is a prevalence of pathological voices compared to healthy ones, the former numbering 150 (52 male and 98 female), the latter 58 (21 male and 37 female).

b. Decomposition of wavelet signals

Our approach relies on the application of discrete wavelet decomposition to shape the feature vector of our speech signal. This decomposition is accomplished by being guided by the specific choice and order of the wavelets used in the calculation of the coefficients. The chunks of the speech signal are fragmented into four levels using the Daubechies wavelet 'db4', creating four details and an approximation for each signal. Here are the steps:

- Step 1: Signal Preparation: We take a one-dimensional signal to analyze.
- Step 2: Filtering by the Low Pass Filter (h): We apply a low pass filter (h) to the signal to extract the low frequency components. The result of the filtering is the approximation signal (cA), which contains the overall characteristics of the signal at this scale.
- Step 3: Filtering by the High Pass Filter (g): We apply a high pass filter (g) to the signal to extract the high frequency components. The result of the filtering is the signal detail (cD), which contains the fine variations and details specific to the signal at that ladder.
- Step 4: Undersampling: We reduce the temporal resolution of the approximation signal (cA) and the detail signal (cD) by retaining each second sample.
  This subsampling reduces the size of the signals but preserves essential characteristics at different scales.
- Step 5: Repeating the Process (Iterations): We repeat steps 2 to 4 for the approximation signal (cA) obtained in the previous step. Each iteration divides the signal into new approximation (cA) and detail (cD) components on a finer scale.
- Step 6: Repeat Until Desired Level: We repeat steps 2 to 5 for a certain predefined number of iterations (decomposition levels) or until the desired temporal resolution is reached.
- Step 7: Approximation and Detail Coefficients: At the end of the process, we obtain approximation (cA) and detail (cD) coefficients at different levels of decomposition.
  The approximation coefficients contain the overall characteristics of the signal at different scales. Detail coefficients contain the fine variations and specific details of the signal at different scales.

c. Extracting parameters

After completing the wavelet transform, we undertook the crucial step of extracting the parameters. This phase is essential to capture the distinctive characteristics of the signals vocal, necessary to distinguish between pathological and normal voices. We calculated a set of parameters that provide information about different properties of the speech signal, thus contributing to the construction of a solid classification model. Before extracting the parameters, we normalized the speech signal to bring all signal values into a specific range, often between -1 and 1. This step is crucial to avoid that the extreme values do not bias subsequent calculations. Normalization is performed as follows:

normalized signal=original signal / max (original signal)

This normalization ensures that all signal amplitudes fall between -1 and 1, making it easier to compare and extract features.

❖ A4 Approximation parameters:

- Energy: The energy of the signal is an indicator of its overall intensity at that frequency scale. It is calculated by summing the squares of the amplitudes of the normalized signal. Energy is an essential characteristic for quantifying signal strength or intensity at this scale.

$$Energy = \sum_{i=1}^{N}(signal\ A4(i))^2$$

- Average: The average of the signal amplitudes reflects its average level and can be related to the overall loudness of the normalized signal. A significantly different average value could indicate a change in signal level.
- Standard Deviation (SD): The standard deviation measures the dispersion of amplitudes around the mean. It provides information on the variability of the normalized signal. A high standard deviation may indicate significant variations

in the signal, while a low standard deviation may indicate greater consistency.

- $SD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(signal\ A4(i)^2 - average)^2}$

- Variance(V): The variance measures the dispersion of the A4 signal values around their mean.

$$V = \frac{1}{N}\sum_{i=1}^{N}(signal\ A4 - average)^2$$

- Average Frequency: The average frequency represents the weighted center frequency of the approximation signal. It provides information on the distribution of frequencies in the normalized signal.

❖ Detail D4 Parameters:

Similar steps were followed to extract the parameters from detail D4:

- Dominant Frequency: The dominant frequency in the signal is the one with the highest amplitude. It captures fine high-frequency variations of the normalized signal.
- Frequency Band Ratio: The frequency band ratio measures the energy distribution between the low and high frequency bands of the signal.
- Spectral entropy: Spectral entropy measures the complexity of the signal in terms of energy distribution over different frequencies.

d. Classification (normal or pathological)

To accomplish our goal of classifying voices into pathological and normal categories, we adopt a machine learning-based approach, specifically using an SVM. This method has proven effective in many classification tasks, including ours. We extracted relevant features from the voice recordings of the database used.

These features will be used as inputs for our SVM model. By adjusting the SVM parameters, we will seek to obtain the best possible separation, thus making it possible to efficiently classify new voices. Our choice of an SVM is based on its simplicity and effectiveness for moderately sized data sets, like the one we handle with VOICED. Although the specific details of preprocessing, feature extraction have already been carried out, we will detail these steps in the next sections.

In addition, we will present our evaluation criteria which will measure the performance of our model. By capitalizing on the discriminative properties of the SVM, our objective is to create a tool for accurate and reliable classification to differentiate between pathological and normal voices. These advances open the way to various potential applications, both in the medical field and in other related fields.

❖ **Measurement parameters**

The confusion matrix is a fundamental tool in classification model evaluation, and it is particularly useful in the context of pathological voice classification. It allows us to quantify the performance of our model by comparing its predictions with the actual values of the voice samples.

This matrix sorts all cases in the model into categories, determining whether the predicted value matched the actual value. All cases in each category are displayed in the matrix [22]. It has four main entries: TP, TN, FP and FN, organized as follows:

Table 1- Confusion Matrix

	Negative Prediction	**Positive prediction**
Negative Label	True Negatives (TN	False Negative (FN)
Positive Label	False Negative (FN)	True Positives (VP)

- Positive Label: The voice sample is indeed pathological.
- Negative Label: The voice sample is indeed normal.
- Positive Prediction: The model predicts that the voice sample is pathological.
- Negative Prediction: The model predicts that the voice sample is normal.
- TP represents the number of samples that actually belong to the positive class

(pathological voices) and that were correctly classified as such by the model.

- TN represents the number of samples that actually belong to the negative class (normal voices) and that were correctly classified as such by the model.
- FP represents the number of samples that actually belong to the negative class (normal voices), but were incorrectly classified as belonging to the positive class (pathological voices) by the model.
- FN represents the number of samples that actually belong to the positive class (pathological voices), but were incorrectly classified as belonging to the negative class (normal voices) by the model.

❖ The sensibility:

Expressed as a true positive rate, sensitivity measures the ability of our model to correctly detect pathological voice samples. It is calculated as follows:

$$Sensibbilty = VP/(VP+FN) \times 100$$

3. Results and Discussion:

Our method was implemented in Matlab and used by a Physionet database. It goes through three stages; Decomposition of wavelet signals, Extracting parameters and Classification.
Decomposition of wavelet signals is made up of seven stages. Figure 3 shows show that the wavelet decomposition up to the fourth level (A4, D4) for vocal signals: normal and pathological gives better frequency resolution.
After, the extraction of the parameters are calculated for the approximation signals A4 and detail D4, both for a pathological voice and for a normal voice. The results obtained are presented in tables 2 and 3.

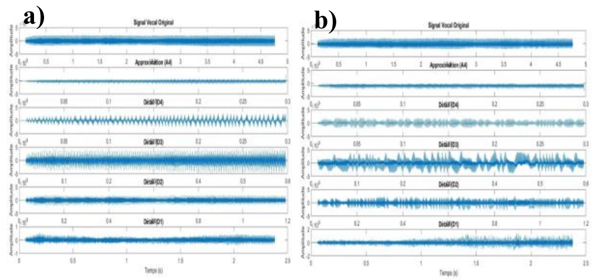

Figure 3- Discrete 4-level wavelet decomposition of : a) normal voice b) pathological voice

Table 2- Parameters of a normal voice signal

Parameters to calculate	A4	Parameters to calculate	D4
Energy	125.1481	Dominant Frequency	1835.9375
Average	-0.0788	Frequency Band Ratio	0.4799
Standard Deviation	0.2151		
Variance	0.0463	Spectral entropy	0.2617
Average frequency	2650.3964		

Table 3- Parameters of a pathological vocal signal

Parameters to calculate	A4	Parameters to calculate	D4
Energy	560.2596	Dominant Frequency	1101.5625
Average	-0.4033	Frequency Band Ratio	0.0213
Standard Deviation	0.2686		
Variance	0.0722	Spectral entropy	0.1851

These parameters, calculated from the details and the normalized approximation, will serve as characteristics for our SVM classification model, aimed at differentiating pathological voices normal voices.

The following figure gives the results of the Confusion Matrix by the application of classifiction by SVM.

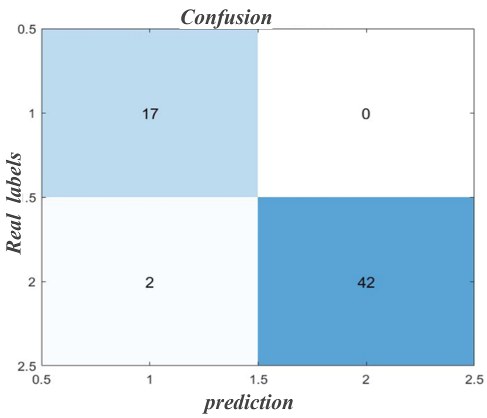

Figure 4 - Results of the confusion matrix

- This number (42) represents the number of pathological voice samples that were correctly identified as pathological by our model. In other words, these are the cases where our model made a positive (pathological) prediction that was correct.
- The number (2) of false negatives indicates the number of pathological voice samples that were misclassified as normal by our model. These samples were actually pathological, but our model misinterpreted them.
- This number (17) represents the number of normal speech samples that were correctly identified as normal by our model. In other words, these are the cases where our model made a negative (normal) prediction that was correct.
- The number (0) of false positives indicates that our model did not make errors in misclassifying normal voice samples as pathological. There are no cases where our model incorrectly predicted a normal voice sample as pathological.

The results of true negative (TN), true positive (TP), false negative (FN), false positive (FP), sensitivity, specificity and classification rate are shown in Table 4:

Table 4 - Classification results obtained using SVM

Inputs	Results
TN	17
FN	2
TP	42
FP	0
Sensitivity	95.45%
Specificity	100%
Classification rate	96.72%

The obtained sensitivity means that our model correctly identified almost 95.45% of all pathological voice samples in our test set. In other words, our model demonstrated a high ability to accurately detect pathological cases among truly pathological samples.

The specificity of 100% indicates that our model correctly identified all normal speech samples in our test set. This means that our model made no errors in identifying normal speech samples, which is very positive.

The classification rate represents the overall accuracy of our model in terms of correct classification. It encompasses both true positives, true negatives, false positives, and false negatives, providing an overview of the overall performance of our model. Our model correctly classified almost 96.72% of all samples, both pathological and normal, in our test set, demonstrating its overall effectiveness.

Table 5 presents a comparative study between our method and the methods of Variations of Pitch+neuron network [18], continuous wavelet transform + SVM [16], discrete wavelet transform + SVM [19], Mel-Frequency Cepstral Coefficients (MFCC) +SVM [20] and MFCC+SVM+ Variations of Pitch [21].

Reading Table 5 by line shows that the classification rate of our method is very high compared to other methods. The results obtained show the effectiveness of the proposed method.

Table 5- Comparative study of our work with other methods

References	Classification rate
[18]	85%
[16]	95.17%
[19]	93.1%
[20]	97%
[21]	95.70%
Our method	96.72%

4. Conclusion

This article was dedicated to the classification of pathological voices, a crucial discipline for vocal health. Through an in-depth exploration of the foundations of the speech signal, a methodology integrating signal processing and classification, as well as the presentation of promising results obtained thanks to an SVM model, we pursued the major objective of improving the diagnosis and treatment of vocal disorders.

Our investigations have highlighted the complexity of the production of human speech and the diversity of sounds it generates, whether sonorous or non-sonorous. We highlighted the importance of understanding vocal alterations and presented methods for identifying vocal disorders, highlighting voice parameters as key elements for the analysis and classification of pathological voices.

The methodology we developed, using the Physionet database and the SVM model, has proven to be effective, as evidenced by the high classification rate of 96.72% that

we obtained. These promising results confirm the relevance of our approach in the detection of vocal pathologies.

However, it is essential to note that our work does not stop here. Prospects for improvement remain, in particular the use of richer and more diversified databases to refine our models. Furthermore, the exploration of advanced machine learning techniques and the integration of artificial intelligence could open new avenues in the field of pathological voice classification.

Ultimately, this work represents a significant contribution to improving the diagnosis and management of voice disorders. It highlights the need for increased attention to vocal health and continued investment in more accurate and accessible diagnostic methods. With a focus on the effectiveness of pathological voice classification, we Let us open up promising perspectives for the care and quality of life of affected individuals.

We hope this work will stimulate further research in the field of vocal health and help provide a brighter future for people with vocal disorders.

5. References:

Malak, A., & Ghulam, M., & Mansour. A. (2014). Detection and classification of voice pathology using feature selection. The 11th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA), Doha, Qatar, 571-577.

Lotfi, S., & Haytham, B., & Adnène. C. (2009). Identification Des Voix Pathologiques Par Réseaux De Neurones. 5th International Conference: Sciences of Electronic. TUNISIA, March 22-26, 1-6.

Nesrine, B., & Amina. B. (2016). Identification des troubles de la voix par l'analyse temps-fréquence et les machines à vecteurs de support. université M'hamed bougara Boumerdes Algérie, projet de fin d'études,1-87.

Hammami, I., & Salhi, L., & Labidi. S. (2020). Voice Pathologies Classification and Detection Using EMD DWT, Analysis Based on Higher Order Statistic Features. University of Tunis El Manar, Higher Institute of Medical Technologies of Tunis, Research Laboratory of Biophysics and Medical Technologies, 1006, Tunis, Tunisia,162-171.

MAROUA, G., & RAHMA. K. (2020). Développement d'un système de détection de la voix pathologique. projet fin étude, l'UNIVERSITÉ BADJI MOKHTAR, ANNABA, 1- 43.

BOUDJELLABA, M., & BOUDJERIDA. N. (2021). Identification des troubles de la voix à l'aide des techniques d'apprentissage automatique. projet fin étude, l'Université Mohammed Seddik Ben Yahia-Jijel, Algérie,1-71.

Godino-Llorente, J. I., & Gómez-Vilda, P., & Sáenz-Lechón, N., &, Blanco-Velasco,M., & CruzRoldán, F., & et al. (2005). Support vector machines applied to the detection of voice disorders," in Nonlinear Analyses and Algorithms for Speech Processing,219-230.

Cesari, U., & De Pietro, G., & Marciano, E., & Niri, C., & Sannino, G.., &, and Verde. A. (2018). A new database of healthy and pathological voices. Computers & Electrical Engineering, vol. 68(5), pp. 310-321.

Authors' Profiles

Dr. Taouli Sidi Ahmed received the B. S. degree in Electronics at the university of Abou Bekr Belkaid of Tlemcen, then his first post-graduation degree in Signal and Systems and his Ph.D. degree in Signal processing from the same university in 2013. Since 2006 he is research professor in a Biomedical Engineering Laboratory. Her research interests are focused on physiological processing signal, biomedical electronics, biomedical modeling and image processing.