Indonesian Journal of Electrical Engineering and Computer Science
Vol. 40, No. 2, November 2025, pp. 640∼653
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v40.i2.pp640-653

Laryngeal pathology detection using EMD-based voice acoustic features analysis and SVM-RBF

Sofiane Cherif[1], Abdelhafid Kaddour[1], Abdelmoudjib Benkada[2], Said Karoui[2], Ouissem Chibani Bahi[1], Asmaa Bouzid Daho[1]
[1]Laboratory of Signals, Systems and Data (LSSD), Department of Electronic, Faculty of Electrical Engineering, University of Sciences and Technology of Oran Mohamed Boudiaf (USTO-MB), Oran, Algeria
[2]Laboratory of Intelligent Systems Research (LARESI), Department of Electronics, Faculty of Electrical Engineering, University of Sciences and Technology of Oran Mohamed Boudiaf (USTO-MB), Oran, Algeria

Article Info

Article history:

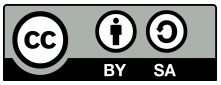
Received Sep 6, 2024
Revised Jul 22, 2025
Accepted Oct 14, 2025

Keywords:

Acoustic features
EMD
Laryngeal pathology
SVM
Voice analysis

ABSTRACT

Traditional techniques for detecting laryngeal pathologies, such as laryngoscopy and endoscopy, are costly and invasive. This study presents a novel approach for detecting laryngeal disorders using empirical mode decomposition (EMD)-based acoustic features analysis and support vector machine (SVM) with a radial basis function (RBF) kernel. The experiments were conducted using the Saarbrücken voice database (SVD). The voice signals were then decomposed using EMD to extract the intrinsic mode functions (IMFs). The IMF with the highest energy value was selected as the most relevant. A set of acoustic features, including mel-frequency cepstral coefficients (MFCCs), linear predictive cepstral coefficients (LPCCs), Pitch (fundamental frequency), higher-order statistics (HOSs), zero-crossing rate (ZCR), spectral centroid (SC), and spectral roll-off (SRO), is derived from the most relevant IMFs and fed into an SVM classifier to differentiate between healthy and pathological voices. Experimental results demonstrate the effectiveness of the proposed methodology, achieving a high classification accuracy of 94.5%, a sensitivity of 94.2%, a specificity of 95.3%, and an F1 score of 96.1%, outperforming conventional approaches. These results highlight the potential of EMD-based voice analysis as a non-invasive and reliable tool for early diagnosis of laryngeal disorders.



Corresponding Author:

Sofiane Cherif
Laboratory of Signals, Systems and Data (LSSD), Department of Electronic, Faculty of Electrical Engineering
University of Sciences and Technology of Oran Mohamed Boudiaf (USTO-MB)
P.O. Box 1505, El Mnaouar, 31000 Oran, Algeria
Email: sofiane.cherif@univ-usto.dz

1. INTRODUCTION

Speech production is a vital function of the vocal tract system, enabling the creation of speech sounds. Impaired voice production can significantly impact an individual's quality of life. Speech pathologists assess impairments affecting communication, language, and voice [1]. The human voice plays a crucial role in facilitating communication and social interaction. However, improper voice use can lead to various problems. Approximately 25% of the world's population suffers from voice disorders [2], which are often caused by conditions affecting the larynx and vocal cords, known as laryngeal pathologies [3]. Conventional diagnostic techniques, such as stroboscopy and laryngoscopy, are commonly used but can cause patients discomfort. Non-invasive methods, such as electroglottography (EGG) and self-assessment, offer alternatives but require

specialist expertise for accurate analysis [4], [5].

To address these challenges and enhance the accuracy of voice disorder detection, researchers have developed various models that extract vocal characteristics, such as mel-frequency cepstral coefficients (MFCCs) and linear predictive cepstral coefficients (LPCCs). These models utilize large voice databases, such as the Saarbrückenvoice database (SVD), and employ advanced classification techniques, including support vector machine (SVM), Gaussian mixture models (GMM), and universal background model Gaussian mixture models (GMM-UBM). Advances in artificial intelligence and machine learning have significantly improved the efficiency of these classification algorithms, enabling more precise and non-invasive detection of laryngeal pathologies [6]. Various innovative approaches, particularly those leveraging deep learning techniques, have achieved significant advancements in voice disorder detection.

Alhussein and Muhammad [7] have developed a system for detecting speech disorders using deep learning techniques. They trained their model on the SVD dataset and evaluated it using the Massachusetts eye and ear infirmary voice disorders database (MEEI). The visual geometry group-16 (VGG16) and CaffeNet algorithms achieved 94.5% and 94.1% accuracy rates, respectively. Leveraging deep convolutional neural networks (CNNs) further improved the accuracy to 97.5%.

Hammami [8] proposed a technique that utilizes wavelet coefficients to classify vocal disorders. Their analysis was based on sustained vowel recordings of the sound /a/ from the SVD dataset. Through experiments with various GMM, they found that incorporating the teager energy operator and using 32 Gaussian mixtures yielded an accuracy of 96.66%. Conversely, when combining three feature vectors, the accuracy dropped to 92.22%.

Fang *et al.* [9] utilized a large set of features, including 430 basic acoustic features (BAFS—basic acoustic features), 84 cepstral coefficients based on the mel S-transform (MSCC—Mel S-transform cepstrum coefficients), and 12 chaotic features. Feature optimization was conducted using radar charts and the F-score, reducing the feature dimensionality from 526 to 96 dimensions for the NKI-CCRT corpus and 104 dimensions for the SVD corpus. These optimized features were fed into an SVM classifier to detect voice disorders. However, their approach achieved only 84.4% accuracy on the NKI-CCRT database and 78.7% on the SVD database. Al-Dhief *et al.* [10] suggested a way to get MFCC features from the SVD database and use them with the OS-LEM (online sequential extreme learning machine) classifier. The approach achieved a maximum accuracy of 91.17%, recall of 91%, F-measure of 87%, G-mean of 87.55%, and specificity of 97.67%.

Ribas *et al.* [11] developed a model based on deep neural networks (DNN) to differentiate between healthy and pathological voices. The model achieved maximum accuracy rates of 80.71% for sentences and 82.8% for vowels (/a/, /i/, /u/). The authors utilized the automatic voice disorder detection (AVDD) system with self-supervised representations to extract distinctive auditory features. They incorporated a feedforward layer with a class-token transformer to consolidate temporal feature sequences. The researchers augmented the training dataset with out-of-scope data to address data availability concerns. Experimental results demonstrated a classification accuracy of 93.36%, representing significant improvements of 4.1% without data augmentation and 15.62% with data augmentation. Using self-supervised (SS) representations in AVDD resulted in an accuracy rate of 90% [11]. Lee [12] employed deep learning techniques to classify voice samples, specifically using feedforward nural networks (FNN) and CNN. Their study found that utilizing the LPCCs, the CNN classifier achieved a maximum accuracy of 82.69% for the vowel /a/ in male subjects.

Ding *et al.* [13] utilized voice signal analysis to develop a method for the early diagnosis and treatment of voice disorders. They also introduced a novel computer-aided assessment approach for pathological voice classification (CS-PVC), specifically designed to distinguish between pathological and healthy voices in areas with significant discrepancies. The model achieved identification accuracy of 81.6% on the SVD dataset and 82.2% on the self-built Shenzhen People's Hospital voice database (SZUPD).

Javanmardi *et al.* [14] conducted a comparative analysis of various data augmentation (DA) techniques for vocal pathology detection, evaluating three temporal methods (noise addition, pitch shifting, and time stretching), one time-frequency technique (SpecAugment), and two vocoder-based approaches (modifying the harmonic-to-noise ratio (HNR) and glottal pulse length). The extracted features include static and dynamic MFCCs, the spectrogram, and the mel-spectrogram, which were then fed into machine learning models (SVM and random forest) and deep learning models (long short-term memory (LSTM) and CNN). The best performance, achieved with a 2D CNN, reached an accuracy of 80% on the SVD database [14].

Albadr *et al.* [15] improved the detection and classification of voice pathologies (VP) using a fast-learning network (FLN) classifier based on MFCCs features. Their study comprised two phases: the first phase

analyzed vocal samples of sustained vowels (/a/, /i/, and /u/) along with spoken phrases. In contrast, the second phase focused on vocal samples from three common voice disorders—paralysis, polyps, and cysts—using the vowel /a/ spoken in a neutral tone. The experimental results achieved an accuracy of 84.64%, a precision of 97.39%, a recall of 86.05%, an F-measure of 86.80%, a G-mean of 86.81%, and a specificity of 88.24%.

According to the literature, traditional methods for identifying laryngeal pathologies rely on vocal signal analysis. However, they have several limitations, particularly the lack of proper pre-processing of voice datasets. Researchers often extract features directly and classify them using a limited number of samples, making it challenging to eliminate residual noise in the reconstructed signal. This leads to oscillations that distort mode decomposition. Additionally, these approaches hinder the systematic evaluation of extracted parameters. To address these issues, we propose a novel method, described in section 2, to improve the detection of laryngeal disorders from speech signals.

This article is structured as follows: section 2 presents the proposed framework, detailing the materials and methodologies used in this study, encompassing both theoretical and practical aspects. Section 3 provides an in-depth discussion of the results, evaluating the effectiveness of the proposed method in detecting laryngeal issues. Finally, section 4 concludes with key findings and suggests potential directions for future research on diagnosing laryngeal pathologies.

2. METHOD

Figure 1 presents the block diagram illustrating the proposed methodology for the accurate and unbiased diagnosis of laryngeal pathologies. This methodology consists of four key steps: silence removal, low-pass filtering, normalization, and empirical mode decomposition (EMD). This method decomposes the vocal signal into IMFs, representing its harmonic components. The most relevant IMFs are selected based on their maximum temporal energy and are framed into short segments (0.1-second duration with 0.01-second overlap) for analysis. Each frame is then multiplied by a Hamming window to minimize discontinuities at the beginning and end of the signal, thereby enhancing the accuracy of the frequency analysis. The Hamming window is the same length as the frame.

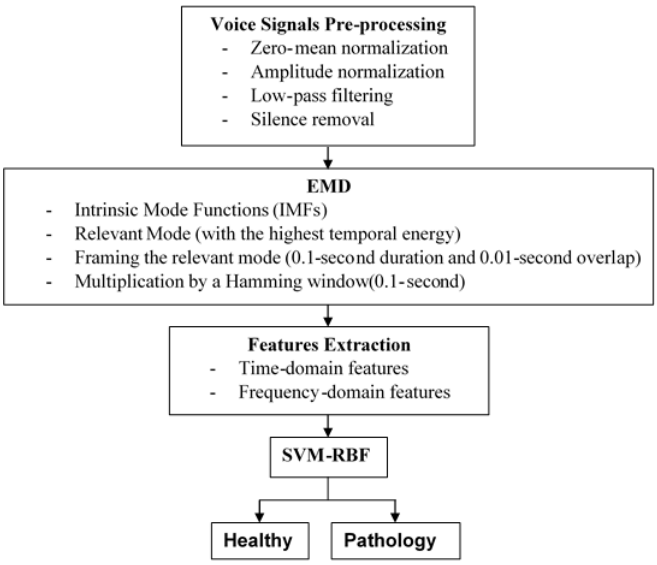


Figure 1. Block diagram illustrating the proposed methodology

Afterward, we extract seven features: Pitch (fundamental frequency), spectral roll-off (SRO), spectral centroid (SC), zero-crossing rate (ZCR), higher-orderstatistics (HOSs), LPCCs, and MFCCs. Finally, each extracted feature serves as input for a support SVM-RBF classifier, enhancing the accuracy of laryngeal pathology diagnosis. The originality of this study lies in integrating voice signal pre-processing and empirical mode decomposition to extract acoustic features. The main contributions of this study are as follows:

- Developing a non-invasive, low-cost method for the detection of laryngeal pathologies
- Experimental validation of the effectiveness of the proposed system using the SVD database
- Using more advanced voice signal pre-processing methods, including different feature extraction and classification algorithms, to make diagnosing laryngeal pathology much more reliable and accurate.

2.1. Database

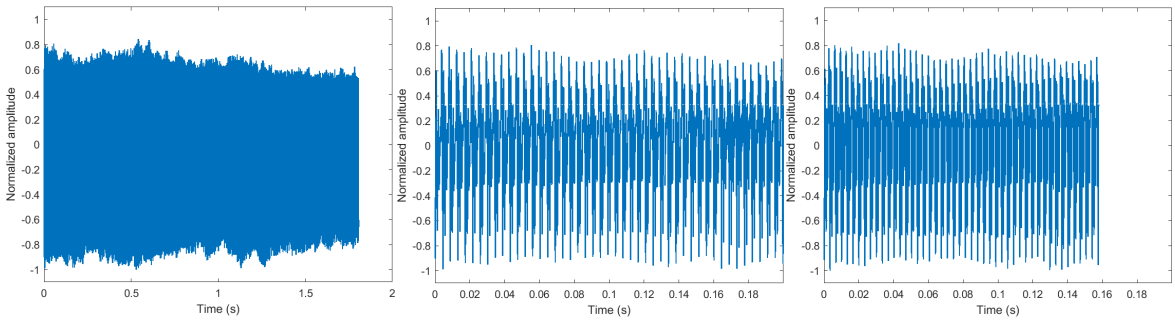
This study utilized the SVD database, an online repository containing over 2,000 audio files featuring three distinct vowel sounds: /a/, /i/, and /u/. Each file has a duration ranging from 1 to 4 seconds and is sampled at a frequency of 50 kHz with a 16-bit resolution. For analysis, we selected vocal signals of the sustained neutral vowel /a/ from a group of 200 healthy males and 91 males with pathological conditions. The pathology subset includes recordings from four specific conditions: 50 cases of laryngitis, 19 cases of vocal cord cancer, 5 cases of Reinke's edema, and 17 cases of vocal cord polyps.

2.2. Vocal signals preprocessing

Before using vocal signals in speech-processing applications, performing pre-processing tasks such as zero-mean normalization, amplified normalization, low-pass filtering, and silence removal is important. Subtracting the mean from a signal centers it around zero, making the average of all the signal samples equal to zero. This process is commonly used to prepare data for machine learning algorithms. The signal is then scaled by dividing each sample by the maximum absolute value. This ensures that the signal's peak is normalized to 1 if the peak is positive or -1 if the peak is negative. We applied a low-pass filter with a cutoff frequency of 1 kHz to isolate the relevant low-frequency components and remove unwanted high-frequency components. Silence removal refers to detecting and removing periods of silence in a signal while maintaining its timing.

This method uses an energy threshold to identify silent periods. In this study, the threshold was set at 2% of the maximum energy level. Any segment with energy below this threshold was considered silent. Vocal signals primarily contain energy at lower frequencies, while non-vocal signals typically have higher frequencies [16]. As illustrated in Figure 2, we present the preprocessing steps applied to the vocal signal of speaker 563 from the SVD database to improve clarity.

Figure 2(a) shows the voice signal 114-a-n.wav after the application of low-pass filtering and normalization. The signal is centered around zero, reflecting the attenuation of high-frequency components and the standardization of the amplitude scale. Figure 2(b) displays a 10,000-sample excerpt of the same signal, corresponding to a duration of 0.2 seconds, to facilitate visual observation. This excerpt allows for a more detailed analysis of the waveform of the preprocessed signal, enabling a localized examination of its acoustic content. Figure 2(c) illustrates the voice signal after silence removal (7,893 samples, corresponding to a duration of 0.1579 seconds). The reduced signal length highlights the effective elimination of silent segments.



(a) (b) (c)

Figure 2. Preprocessing of the voice signal: (a) low-pass filtered and normalized voice signal 114-a_n.wav, (b) 10,000-sample excerpt of a low-pass filtered and normalized voice signal, and (c) voice signal after silence removal (7,893 samples corresponding to 0.1579-second duration)

2.3. Empirical mode decomposition

Many researchers have used EMD to process vocal signals due to its excellent performance with this specific type of signal [17]-[20]. To detect the presence of voice in a non-stationary speech signal, we applied EMD to decompose it into a sequence of oscillatory patterns known as IMFs and a residual component, as shown in (1).

$$x(n) = r_k(n) + \sum_{i=1}^{k} IMF_i(n) \tag{1}$$

Where $x(n)$ is the digitized voice signal, n representing the sample, k is the number of IMFs extracted and $r_k(n)$ is the residual.

We incorporated the stopping condition proposed by Huang *et al.* [17] for the sifting procedure. This criterion limits the standard deviation (SD) between two consecutive sifting results typically between 0.2 and 0.3. For an IMF to be considered genuine, it must satisfy two criteria: the difference between the number of zero crossings and the number of extrema must not exceed one, and the average value of the envelope formed by the local maxima and minima must be zero. Figure 3 illustrates the decomposition process as well as the criteria used to identify the most relevant IMFs, summarizing the key steps of our method. It highlights both the decomposition procedure and the steps used to extract acoustic information from the most significant components.

The IMFs, shown in Figure 3(a), are obtained through an iterative sifting process, which involves the following steps:

i) Determine all extrema (local maxima and minima) of the signal $x(t)$.

ii) Estimate the values of the minima and maxima using cubic spline interpolation, creating the lower envelope $e_{\min}(t)$ and the upper envelope $e_{\max}(t)$.

iii) Determine the envelope's mean by applying the following formula:

$$m_1(t) = \frac{e_{\max}(t) + e_{\min}(t)}{2} \tag{2}$$

iv) Calculate the IMF by calculating the difference between the $x(t)$ and $m_1(t)$ signals.

$$x(t) - m_1(t) = h_1(t) \tag{3}$$

v) If $h_1(t)$ is an IMF, it is defined as the first IMF component of $x(t)$. Alternatively, $h_1(t)$ is considered the original signal.

vi) Iterate the preceding steps, treating $h_1(t)$ as the new $x(t)$, and obtain $h_{11}(t)$. If $h_{11}(t)$ is an IMF, stop the process. Otherwise, continue iterating.

After the decomposition, we have identified the IMF with the highest energy value as the most relevant IMFs. The energy is calculated using (4).

$$E_k = \sum_{n=1}^{N} [IMF_k(n)]^2 \tag{4}$$

Where E_k is the energy of the k–th IMF, N is the length of the backscattered signal, and $IMF_k(n)$ is the value of the k–th IMF at sample n.

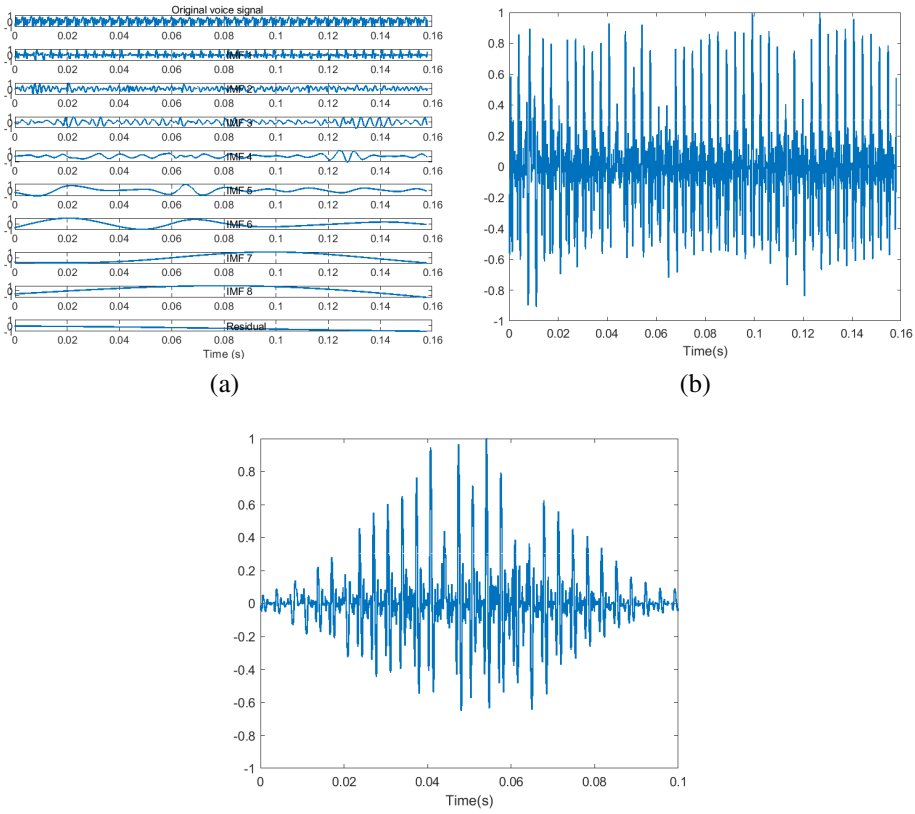
The relevant IMF obtained (Figure 3(b)) is segmented into 0.1-second intervals and then multiplied by a Hamming window of the same length (Figure 3(c)) to extract acoustic features.

2.4. Feature extraction

2.4.1. Mel-frequency cepstral coefficients

The MFCCs are extensively utilized features in speech and audio processing. It denotes the short-term power spectrum of an auditory input, emulating human speech perception. MFCCs are crucial for identifying vocal abnormalities in the vocal domain [21], [22]. Figure 4 illustrates the steps involved in computing MFCCs. The pre-emphasis step enhances high frequencies to balance the spectrum. The fast fourier transform (FFT) then converts the time-domain signal into a frequency spectrum. Subsequently, a Mel filter bank is applied

to map frequencies onto the mel scale, which aligns with human auditory perception. Finally, the amplitudes are converted to a logarithmic scale (similar to human perception) and subjected to a discrete cosine transform (DCT), extracting the most relevant MFCCs for classifying laryngeal diseases.



(a) (b)

(c)

Figure 3. Decomposition of voice signal: (a) IMFs, (b) the relevant mode, and (c) relevant mode multiplied by the hamming window (0.1-second)

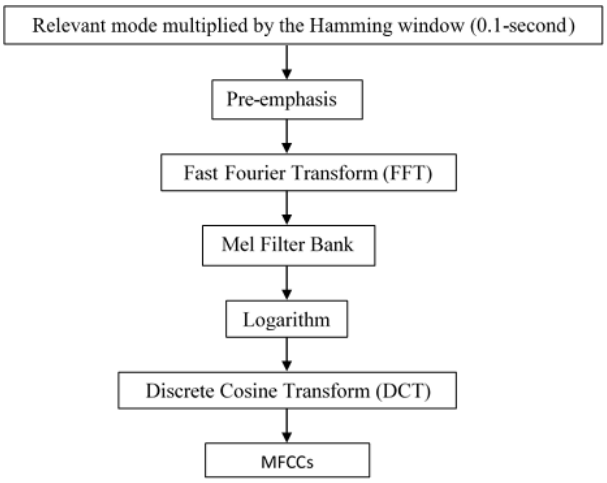


Figure 4. Steps to compute MFCCs

2.4.2. Linear predictive cepstral coefficients

LPCCs are an advanced signal processing technique used to estimate the source signal of vocal sounds. This method utilizes LPCCs—also referred to as CPLC—to perform a detailed analysis of the vocal signal. The primary goal of LPCCs is to model the signal's spectral envelope to extract its essential features. The vocal tract is an infinite impulse response (IIR) filter modeled through a recursive and graphical approach [23]. This modeling process is described in (5).

$$H(z) = \frac{G}{1 + \sum_{k=1}^{p} a_p(k) Z^{-k}} \tag{5}$$

Where p is the number of poles, G denotes the filter gain, and $a_p(k)$ are the coefficients.

The extraction of LPCCs involves a series of sequential steps, as illustrated in Figure 5. First, the relevant signal segment—multiplied by a 0.1-second hamming window—is modeled using a linear predictive model, which assumes that the current sample can be estimated as a linear combination of previous samples. The model coefficients are obtained by minimizing the prediction error. The autocorrelation function of the predicted signal is then computed to assess the similarity between different parts of the signal. Subsequently, the iterative Levinson-Durbin algorithm is employed to derive the LPCCs from the autocorrelation function. Finally, the LPCCs are transformed into the cepstral domain by applying the discrete cosine transform (DCT) [24], [25].

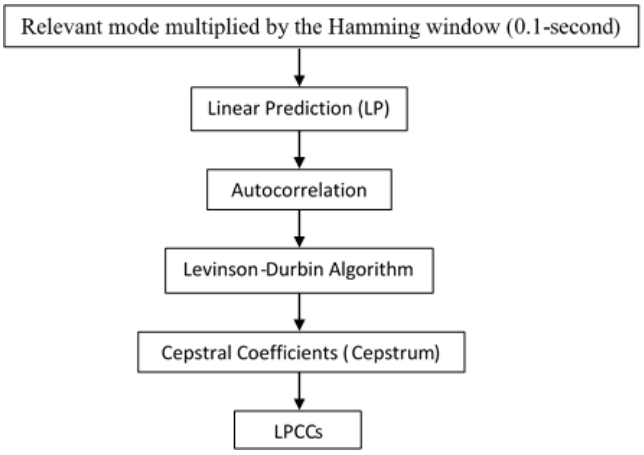


Figure 5. Steps to compute LPCCs

2.4.3. Pitch

The fundamental frequency (F_0), often called pitch, is the frequency at which the vocal cords vibrate when producing voiced sounds. This frequency is a crucial indicator of laryngeal diseases. Several methods for calculating F_0 are described in the literature, including those based on autocorrelation, spectral analysis, and combinations of these techniques [20]. For our study, we chose the autocorrelation method, as defined by (6).

$$R[k] = \sum_{n=0}^{N-k-1} x[n] \cdot x[n+k] \tag{6}$$

Where:

- $R[k]$ represents the one-lag autocorrelation function k,
- $x[n]$ is the input signal at time n,
- k denotes the shift index (lag),
- N is the length of the signal.

The first peak (local maximum) in the autocorrelation function, after the peak at $k = 0$, corresponds to the fundamental period of the signal. The period T_0 is the distance between this peak and $k = 0$.

2.4.4. Higher order statistics

Our work explicitly examined the HOSs characteristics, focusing on the third-order moments (skewness) and fourth-order moments (Kurtosis). One notable benefit of these HOSs features is their compatibility with periodic and non-periodic signals. Skewness quantifies the lack of symmetry in a voice's probability distribution, whereas Kurtosis measures the extent to which a distribution is flat and contains impulsive elements in a signal. These two statistics provide a valuable method for analyzing voice features and diagnosing pathology laryngeal, assessing data distribution, and identifying impulsive components. We compute the Skewness and Kurtosis using (7) and (8) in sequential order [26]-[28]:

$$\gamma_3 = \frac{\sum_{n=1}^{N}(x_n - \mu)^3}{(N-1)\sigma^3} \tag{7}$$

$$\gamma_4 = \frac{\sum_{n=1}^{N}(x_n - \mu)^4}{(N-1)\sigma^4} \tag{8}$$

Where γ_3 and γ_4 denote the measures of skewness and Kurtosis, respectively, N the number of samples, μ the mean and σ the SD.

2.4.5. Zero-crossing rate

The ZCR is a quantitative measure employed to assess the frequency characteristics of a signal. The term "sign change rate" refers to the frequency at which a signal changes its polarity within a specific time frame. More precisely, it counts the number of times the signal changes from positive to negative values (or vice versa) and then standardizes this tally by dividing it by the total duration of the frame. The following mathematical expression determines the zero-crossing rate:

$$Z_n = \frac{1}{w_l}\sum_{m=1}^{w_l} |\text{sgn}[x_n(m)] - \text{sgn}[x_n(m-1)]| \tag{9}$$

The length of the frame is represented by w_l, the frame number is represented by m, and the sign function is represented by sgn.

$$\text{sgn}[x_n(m)] = \begin{cases} 1 & \text{si } x_n(m) > 0, \\ 0 & \text{si } x_n(m) = 0, \\ -1 & \text{si } x_n(m) < 0. \end{cases} \tag{10}$$

2.4.6. Spectral centroid

The spectral centroid is a crucial feature used to identify voice disorders. It represents the "center of gravity" of the spectrum and is computed using frequency and amplitude information derived from the fourier transform [29], [30]. The spectral centroid indicates the frequency in Hertz (Hz) at which the spectral energy is balanced or evenly distributed. It is calculated as the weighted average of the frequencies contained in the signal, as expressed by (11).

$$\text{Spectral centroid} = \frac{\sum_{k=1}^{N} f_k \cdot S_k}{\sum_{k=1}^{N} S_k} \tag{11}$$

Where N represents the number of spectral bins or frequencies, f_k is the frequency of the k-th spectral bin, and S_k denotes the the amplitude of the k-th spectral bin.

2.4.7. Spectral roll-off

The term "spectral roll-off" refers to a metric that is used to define a filter that is intended to decrease the amplitude of frequencies that fall outside of a particular range. This technique is frequently used to reduce undesired frequencies in a transmission. It is a measure that identifies the frequency at which a specific percentage of the total energy in a spectrum is concentrated below. The equation for SRO states that the spectral

energy accumulated up to the i-th bin is proportional to the total energy contained between the b_1 and b_2 bins and it is typically expressed as follows [28]:

$$\text{Roll-off spectral}(i) = \sum_{k=b_1}^{i} S_k = K \sum_{k=b_1}^{b_2} S_k \tag{12}$$

where S_k represents the spectral amplitude at the k frequency bin. $b_1$and b_2 are the band edges over which the spectral spread is calculated, and K represents the percentage of total energy. The equation expresses that the spectral energy accumulated up to the ii-th bin is proportional to the total energy contained between the b_1 and b_2 bins.

2.5. Classification

Several techniques are available for classifying laryngeal disorders based on vocal signals, including CNNs, AlexNet, SVMs, random forests, K-nearest neighbors (KNN), decision trees, and deep neural networks (DNNs). Each algorithm offers distinct advantages, improving classification accuracy depending on the context and dataset [4], [31].

In our study, we selected a SVM with a RBF kernel. The SVM-RBF is a supervised learning model designed to construct an optimal hyperplane that separates data into two distinct classes. One of its key strengths lies in its deterministic nature, as it does not rely on probabilistic assumptions. Such an approach can lead to more consistent and interpretable results in specific applications.

The SVM-RBF's goal is to find the hyperplane that maximizes the margin—the distance between the hyperplane and the closest support vectors. This margin serves as a decision boundary that best differentiates the two classes. A wider margin typically improves the model's generalization capability, enabling it to more accurately classify new, unseen data. Additionally, the margin-based approach contributes to robustness by reducing the model's sensitivity to outliers and noise in the dataset [27].

To optimize the performance of the SVM-RBF model for our specific dataset, we conducted an exhaustive parameter search. In particular, we fine-tuned two crucial parameters: the kernel scale (γ) and the box constraint (C). The kernel scale regulates the impact of individual training samples on the configuration of the decision border, whereas the box constraint mediates the balance between optimizing the margin and reducing classification mistakes [32]. By carefully adjusting these parameters, we could regulate the complexity of the decision surface and enhance the model's effectiveness in classifying vocal signals associated with laryngeal disorders. The RBF kernel used in SVMs is mathematically defined as follows:

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|} \tag{13}$$

where:

- x_i and x_j are feature vectors in the input space,
- $K(x_i, x_j)$ is the kernel function that computes the similarity between two data points x_i and x_j,
- $\|x_i - x_j\|$ represents the Euclidean distance between the two data points x_i and x_j,
- γ is a parameter that controls the spread of the kernel. A higher value of γ results in a narrower kernel, meaning that only points that are very close to each other will be considered similar. Conversely, a lower value of γ makes the kernel wider, considering more distant points as similar.

While the box constraint (C) is a regularization parameter that controls the trade-off between achieving a low training error and maintaining a simpler decision boundary. A higher value of C penalizes misclassifications more heavily, leading to a complex decision boundary that may overfit the data, whereas a lower C allows for more classification errors, promoting a simpler and more generalized model.

$$\min_{\omega, b, \epsilon} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{l} \epsilon_i \tag{14}$$

Subject to the constraints:

$$y_i(w^T x_i + b) \geq 1 - \epsilon_i, \quad \epsilon_i \geq 0, \quad i = 1, \ldots, l \tag{15}$$

where w denotes the normal vector defining the hyperplane, b represents the bias shifting the hyperplane, l is the total number of data points, ϵ are the slack variables allowing for tolerance of classification errors, and

$y_i \in \{+1, -1\}$ is the class of the sample x_i. We investigated the optimization parameters $C = 2^k$ and $\gamma = 2^m$, where k and m are integers chosen within the range of -20 to 20. By fine-tuning these parameters, we aim to enhance classification performance while preserving a balance between accuracy and generalization.

We aim to enhance classification performance by fine-tuning these parameters while preserving a balance between accuracy and generalization. We evaluated these automated classification and detection methods for laryngeal diseases using four key metrics: accuracy, sensitivity, specificity, and the F1 score. In this case, the algorithm classifies samples as either pathological or healthy, accordingly labeling them as true positives (TP) or false negatives (FN). Conversely, healthy samples are classified as either pathological or healthy, corresponding to true negatives (TN) and false positives (FP). The following equations define a variety of these performance measures.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

$$\text{Sensitivity (Recall)} = \frac{TP}{TP + FN} \tag{17}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{18}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{19}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{20}$$

3. RESULTS AND DISCUSSION

The proposed laryngeal disease detection and classification method was evaluated using the SVD database, described in section 2.1. In our experiments, 80% of the data was used for training, while 20% was reserved for testing and validation to evaluate the model's performance. Interpreting the confusion matrix is essential for evaluating the model's performance in accurately classifying the different categories (normal or pathological). This evaluation is guided by the metrics defined in section 2.5, which provide a quantitative classification performance assessment. Table 1 presents the metric values corresponding to each feature: MFCCs, LPCCs, HOSs, Pitch, SRO, ZCR, and SC.

Table 1. Evaluation metrics table of different characterization parameters

Parameter	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1 (%)	AUC (%)
14 MFCCs	94.5	94.2	95.3	96.1	94.5
14 LPCCs	85.8	88.7	78.1	88.7	85.5
HOSs	86.1	91.3	71.9	91.3	86.1
Pitch	86.6	87.9	83.5	87.9	89.1
SRO	86.1	86.9	83.9	86.9	86.1
ZCR	79.2	90.6	50.7	90.6	79.2
SC	86.0	93.2	68.2	93.2	86

The metrics presented in Table 1 provide valuable insights into the contribution of each acoustic feature to the classification of normal and pathological voices. Among all the parameters, MFCCs, and LPCCs exhibit the highest performance across all evaluation metrics, indicating their strong discriminative power in detecting vocal pathologies. This result is consistent with previous studies, which highlight the efficiency of cepstral features in capturing relevant information from speech signals. HOSs also show promising results, suggesting that the voice signal's nonlinear characteristics contain useful diagnostic cues. Pitch and SRO demonstrate moderate classification performance, likely because they capture complementary aspects of vocal signal variability that may not be as robust across all samples.

In contrast, the ZCR and SC yield relatively lower metric scores. While these features are useful for capturing general spectral characteristics, their limited ability to capture pathological anomalies may explain their lower impact on classification accuracy. Overall, the results suggest that combining multiple features—particularly cepstral and statistical descriptors—can enhance the model's ability to distinguish between normal and pathological speech patterns.

Figure 6 illustrates each parameter's ROC (receiver operating characteristic) curve, enabling us to calculate the area under the curve (AUC) values. These numbers offer insights into the diagnostic precision and predictive capacity. The performance of the pathological classification was evaluated using the ROC curves' AUC (area under the curve). In the Matlab 2020a environment, the perfcurve() function was used to calculate the AUC, which employs a trapezoidal approximation to determine the area. This allowed for comparison between the extracted parameters.

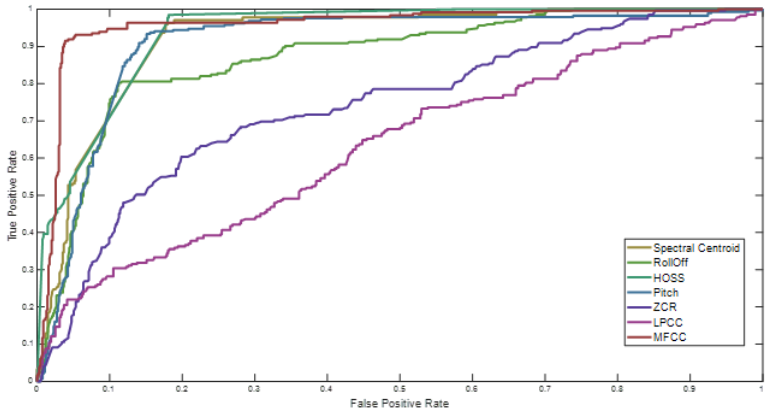


Figure 6. ROC curve analysis of the different models for the classification task, highlighting the SVM as the most frequently used method

The SVM-RBF addresses the classification task by constructing a hyperplane that maximizes the margin between the two classes [33]. To assess the effectiveness of various features - MFCCs, LPCCs, HOSs, Pitch, SRO, ZCR, and SC in classifying speech signals using the SVM-RBF algorithm, we evaluated several performance metrics: accuracy, sensitivity, specificity, and the area under the ROC AUC. These features encompass the speech signal's spectral and temporal characteristics, capturing essential information related to voice quality, frequency content, and dynamic variations in the phonation process.

The results of our analysis show that performance varies depending on the parameters used for classification. The MFCCs features proved to be highly significant in evaluating pathological laryngeal voices, achieving the highest accuracy (94.5%), sensitivity (94.2%), specificity (95.3%), and the largest AUC (94.5%). These results suggest that MFCCs are particularly effective at distinguishing between normal and pathological voice signals. In contrast, the accuracy for the other parameters (LPCCs, HOSs, Pitch, SRO, ZCR, and SC ranged from 79.2% to 86.6%, specificity ranged from 50.7% to 83.9%, and sensitivity ranged from 86.9% to 93.2%. All of these parameters had an AUC greater than 0.5. These findings highlight the varying performance of different criteria and underscore the importance of carefully selecting acoustic features for laryngeal disease classification. Although MFCCs demonstrated excellent sensitivity, specificity, and F1 score, it is important to recognize that each parameter provides unique insights into voice characteristics. The variability observed indicates that a well-chosen combination of parameters could lead to even higher classification accuracy.

Based on the results obtained, we performed a comparative evaluation. For example, Table 2 presents a comparison between our method and other existing approaches using MFCCs features on the same SVD dataset. As discussed in section 1, various strategies have been presented in the literature. The authors of these studies evaluated all methods using accuracy, ensuring consistency with the evaluation criteria in this study. However, our proposed method outperforms these leading approaches, achieving an accuracy rate of 94.5%, demonstrating the effectiveness of the laryngeal disease detection strategy. This study emphasizes the critical role of selecting appropriate acoustic features in diagnosing laryngeal issues and presents a comprehensive

system for classifying these disorders. Addressing these factors will contribute to developing more accurate and reliable diagnostic tools for laryngeal diseases.

Table 2. Comparison of accuracies of our proposed method versus those in the literature

Reference	Features used	Classifier/approach	Accuracy
Fang *et al.* [9]	MFCCs + suprasegmental features	SVM	78.7%
AL-Dhief *et al.* [10]	MFCCs, jitter, shimmer	OS-ELM	91.17%
Lee [12]	MFCCs + spectrograms	CNN, BiLSTM	Up to 82.69%
Ding *et al.* [13]	MFCCs + attention modules	ResNet with attention	81.6%
Javanmardi *et al.* [14]	MFCCs (also PLP, log-mel)	SVM, CNN	Up to 80%
Albadr *et al.* [15]	MFCCs + acoustic parameters	FLN	84.64%
Our method	MFCCs	SVM	94.5%

4. CONCLUSION

This paper explores the use of new features derived from empirical decomposition to assess their effectiveness in identifying laryngeal disorders. We compare these features using a machine classifier with SVM, incorporating optimized hyperparameters such as the kernel scale (gamma) and the box constraint (C). The study evaluates the performance of several parameters in classifying laryngeal diseases. The SVM-RBF classifier is employed to categorize the disorders, while the EMD is used to analyze voice signals and extract the IMFs based on their energy levels. We then derived multiple parameters from these IMFs to serve as inputs for the SVM-RBF algorithm. The results showed that the MFCCs were highly effective in detecting vocal pathologies, achieving optimal levels of precision, sensitivity, specificity, and the highest area under the ROC curve. However, this study has some limitations, primarily the size of the dataset, the potential variability in voice recordings, and the need for further improvements in voice signal processing methods. To address these challenges, our future research will focus on employing improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) to optimize the selection of the IMFs and enhance the efficiency of processing techniques.

Additionally, incorporating adaptive split spectrum processing (ASSP) and quadratic time-frequency distributions (QTFD) could enable a more detailed analysis of time-frequency information, including spectrograms, entropies, and scalograms. Furthermore, exploring deep learning models in combination with hybrid feature selection methods, such as the max-relevance and min-redundancy (mRMR) algorithm, could enhance the accuracy of extracted features and improve classifier performance. Expanding the database by including a larger number of cases representing various types of laryngeal diseases could enhance the model's ability to generalize. Additionally, integrating these techniques and algorithms into a mobile application would greatly simplify the detection of laryngeal diseases for otorhinolaryngology specialists, contributing to the development of a reliable diagnostic tool for early disease detection. The results of the current research are the basis for the development of future techniques.

FUNDING INFORMATION

Authors state no funding involved.

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] V. Srinivasan, V. Ramalingam, and P. Arulmozhi, "Artificial neural network based pathological voice classification using MFCC features," *International Journal of Science, Environment and Technology*, vol. 3, no. 1, pp. 291–302, 2014.

[2] A. Al-Nasheri *et al.*, "Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions," *IEEE Access*, vol. 6, pp. 6961–6974, 2017, doi: 10.1109/ACCESS.2017.2696056.

[3] F. T. AL-Dhief *et al.*, "Voice pathology detection using machine learning technique," in *2020 IEEE 5th International Symposium on Telecommunication Technologies (ISTT)*, Nov. 2020, pp. 99–104, doi: 10.1109/ISTT50966.2020.9279346.

[4] G. Muhammad and M. Alhussein, "Convergence of artificial intelligence and internet of things in smart healthcare: a case study of voice pathology detection," *IEEE Access*, vol. 9, pp. 89198–89209, 2021, doi: 10.1109/ACCESS.2021.3090317.

[5] I. Hammami, L. Salhi, and S. Labidi, "Voice pathologies classification and detection using EMD-DWT analysis based on higher order statistic features," *IRBM*, vol. 41, no. 3, pp. 161–171, Jun. 2020, doi: 10.1016/j.irbm.2019.11.004.

[6] R. Islam, M. Tarique, and E. Abdel-Raheem, "A survey on signal processing based pathological voice detection techniques," *IEEE Access*, vol. 8, pp. 66749–66776, 2020, doi: 10.1109/ACCESS.2020.2985280.

[7] M. Alhussein and G. Muhammad, "Voice pathology detection using deep learning on mobile healthcare framework," *IEEE Access*, vol. 6, pp. 41034–41041, 2018, doi: 10.1109/ACCESS.2018.2856238.

[8] İ. Hammami, "Classification of psychogenic and laryngeal voice diseases based on teager energy operator," *International Journal of Applied Mathematics Electronics and Computers*, vol. 7, no. 3, pp. 49–55, Sep. 2019, doi: 10.18100/ijamec.458230.

[9] C. Fang, H. Li, L. Ma, and M. Zhang, "Intelligibility evaluation of pathological speech through multigranularity feature extraction and optimization," *Computational and Mathematical Methods in Medicine*, vol. 2017, pp. 1–8, 2017, doi: 10.1155/2017/2431573.

[10] F. T. Al-Dhief *et al.*, "Voice pathology detection and classification by adopting online sequential extreme learning machine," *IEEE Access*, vol. 9, pp. 77293–77306, 2021, doi: 10.1109/ACCESS.2021.3082565.

[11] D. Ribas, M. A. Pastor, A. Miguel, D. Martinez, A. Ortega, and E. Lleida, "Automatic voice disorder detection using self-supervised representations," *IEEE Access*, vol. 11, pp. 14915–14927, 2023, doi: 10.1109/ACCESS.2023.3243986.

[12] J.-Y. Lee, "Experimental evaluation of deep learning methods for an intelligent pathological voice detection system using the saarbruecken voice database," *Applied Sciences*, vol. 11, no. 15, p. 7149, Aug. 2021, doi: 10.3390/app11157149.

[13] H. Ding, Z. Gu, P. Dai, Z. Zhou, L. Wang, and X. Wu, "Deep connected attention (DCA) ResNet for robust voice pathology detection and classification," *Biomedical Signal Processing and Control*, vol. 70, p. 102973, Sep. 2021, doi: 10.1016/j.bspc.2021.102973.

[14] F. Javanmardi, S. R. Kadiri, and P. Alku, "A comparison of data augmentation methods in voice pathology detection," *Computer Speech & Language*, vol. 83, p. 101552, Jan. 2024, doi: 10.1016/j.csl.2023.101552.

[15] M. A. A. Albadr *et al.*, "Fast learning network algorithm for voice pathology detection and classification," *Multimedia Tools and Applications*, vol. 84, no. 17, pp. 18567–18598, Jul. 2024, doi: 10.1007/s11042-024-19788-3.

[16] Y. A. Ibrahim, J. C. Odiketa, and T. S. Ibiyemi, "Preprocessing technique in automatic speech recognition for human computer interaction: an overview," *Anale. Seria Informatică*, vol. 15, pp. 186–191, 2017.

[17] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, Mar. 1998, doi: 10.1098/rspa.1998.0193.

[18] N. E. Huang *et al.*, "A confidence limit for the empirical mode decomposition and Hilbert spectral analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 459, no. 2037, pp. 2317–2345, Sep. 2003, doi: 10.1098/rspa.2003.1123.

[19] Zhaohua Wu and N. E. Huang, "Ensemble Empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2011, doi: 10.1142/S1793536909000038.

[20] N. E. Huang and S. S. P. Shen, Hilbert–Huang transform and its applications, vol. 16. *WORLD SCIENTIFIC*, 2014.

[21] K. Bhattarai, P. W. C. Prasad, A. Alsadoon, L. Pham, and A. Elchouemi, "Experiments on the MFCC application in speaker recognition using Matlab," in *2017 Seventh International Conference on Information Science and Technology (ICIST)*, Apr. 2017, pp. 32–37, doi: 10.1109/ICIST.2017.7926796.

[22] A. N. Omeroglu, H. M. A. Mohammed, and E. A. Oral, "Multi-modal voice pathology detection architecture based on deep and handcrafted feature fusion," *Engineering Science and Technology, an International Journal*, vol. 36, p. 101148, Dec. 2022, doi: 10.1016/j.jestch.2022.101148.

[23] S. M. Ali and P. T. Karule, "Development of automation system for disease disorder diagnosis using artificial neural networks and support vector machine," *Journal on Science Engineering & Technology*, vol. 2, no. 01, pp. 103–112, 2015.

[24] I. Hariga *et al.*, "Chronic Laryngitis: Diagnosis and therapeutic approach *(in France: Laryngite chronique: approche diagnostique et therapeutique)*," *Journal Tunisien d'ORL et de Chirurgie Cervico-Faciale*, vol. 30, 2013, doi: 10.4314/jtdorl.v30i1.

[25] S. Jothilakshmi, "Automatic system to detect the type of voice pathology," *Applied Soft Computing*, vol. 21, pp. 244–249, Aug. 2014, doi: 10.1016/j.asoc.2014.03.036.

[26] D. Yilmaz and H. Ankişhan, "Vokal Kist Probleminin Yüksek Dereceli Momentler ve Destek Vektör Makineleri Kullanılarak Tespiti," *Academic Platform Journal of Engineering and Science*, pp. 1–1, Sep. 2018, doi: 10.21541/apjes.380271.

[27] J. Y. Lee and M. Hahn, "Automatic assessment of pathological voice quality using higher-order statistics in the LPC residual domain," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 1, p. 748207, Dec. 2010, doi: 10.1155/2009/748207.

[28] F. Abakarim and A. Abenaou, "Voice pathology detection using the adaptive orthogonal transform method, SVM and MLP.," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 17, no. 14, pp. 90–102, Dec. 2021, doi: 10.3991/ijoe.v17i14.26701.

[29] X. Xu, J. Cai, N. Yu, Y. Yang, and X. Li, "Effect of loudness and spectral centroid on the music masking of low frequency noise from road traffic," *Applied Acoustics*, vol. 166, p. 107343, Sep. 2020, doi: 10.1016/j.apacoust.2020.107343.

[30] A. Lauraitis, R. Maskeliunas, R. Damasevicius, and T. Krilavicius, "Detection of speech impairments using cepstrum, auditory spectrogram and wavelet time scattering domain features," *IEEE Access*, vol. 8, pp. 96162–96172, 2020, doi: 10.1109/ACCESS.2020.2995737.

[31] M. Yalsavar, P. Karimaghaee, A. Sheikh-Akbari, M.-H. Khooban, J. Dehmeshki, and S. Al-Majeed, "Kernel parameter optimization for support vector machine based on sliding mode control," *IEEE Access*, vol. 10, pp. 17003–17017, 2022, doi: 10.1109/ACCESS.2022.3150001.

[32] B. Kumar, O. P. Vyas, and R. Vyas, "A comprehensive review on the variants of support vector machines," *Modern Physics Letters B*, vol. 33, no. 25, p. 1950303, Sep. 2019, doi: 10.1142/S0217984919503032.

[33] H. Kim *et al.*, "Convolutional neural network classifies pathological voice change in laryngeal cancer with high accuracy," *Journal of Clinical Medicine*, vol. 9, no. 11, p. 3415, Oct. 2020, doi: 10.3390/jcm9113415.

BIOGRAPHIES OF AUTHORS

Sofiane Cherif was born on 22.01.1993 in Mohammadia – Mascara, Algeria. He obtained his Master's degree in biomedical instrumentation from the University Aboubekr Belkaid of Tlemcen, Algeria in 2017. Currently, he has been preparing his doctoral thesis in biomedical engineering since April 2022 at the Department of Electronics, University of Science and Technology of Oran Mohamed Boudiaf (USTO-MB), Algeria. He can be contacted at email: sofiane.cherif@univ-usto.dz.

Abdelhafid Kaddour was born in Tlemcen, Algeria. He received his Dipl. El. Ing. degree in electronics, Magister Degree, and Doctorat d'Etat in electroacoustics and signal processing from the University of Sciences and Technology of Oran (USTO-MB), Algeria. He was an Assistant Professor at the Department of Electronics from 1982 where he became an associate from 2004 and Professor from 2009. His current research interests include electronics, acoustics, electro-acoustics, applied ultrasound measurement, biomedical engineering, biomedical signal processing and biological modeling and simulation. He can be contacted at email: kaddour.abdelhafid@gmail.com.

Abdelmoudjib Benkada received his electronic engineering and magister degrees from the University of Science and Technology of Oran Mohamed Boudiaf (USTO-MB), Algeria, in 1999 and 2003, respectively. In 2019, he obtained his Ph.D. in advanced control and vibrations. Since 2005, he has been actively engaged as a teaching and research professional at the Department of Electronics (USTO-MB). His research interests include medical engineering, human acoustics, and nonlinear modeling. He can be contacted at email: abdelmoudjib.benkada@gmail.com.

Said Karoui received a Bachelor's degree in electrical engineering from the University of Sciences and Technology of Oran (USTO-MB), Algeria in 1988, and a Ph.D. degree from Grenoble National Institute of Technology (INPG), France in 1993. He subsequently held various R & D engineering positions in the aerospace, automotive, and telecommunications industries in the Montreal area for 10 years. Since 2005, he has been an Associate Professor at USTO-MB. His research interests include circuits and system testing, design for EMI, design and reliability studies of space applications, speech processing, and the design of embedded digital systems. He can be contacted at email: said.karoui@univ-usto.dz.

Ouissem Chibani Bahi obtained a Master's degree in biomedical instrumentation at the University of Science and Technology of Oran Mohamed Boudiaf (USTO-MB), Algeria, in 2019. In 2024, she completed a second Master's degree in signal and image processing, technologies and imaging course for medicine, at Clermont Auvergne University - University School of Physics and Engineering, Clermont -Ferrand, France. She is currently preparing for a Doctorate in biomedical instrumentation, continuing the research started during her first master's degree. His research interests include non-invasive diagnostics, particularly quantitative ultrasound, to advance this field at the interface of engineering and medicine. She can be contacted at email: ouissem.chibanibahi@univ-usto.dz.

Asmaa Bouzid Daho obtained a Master's degree in biomedical instrumentation at the University of Science and Technology of Oran Mohammed Boudiaf, Algerie, in 2019. Currently has been preparing her doctoral thesis in biomedical instrumentation since April 2022 at the Department of Electronics, University of Science and Technology of Oran Mohamed Boudiaf (USTO-MB), Algeria. She can be contacted at Email: asmaa.bouziddaho@univ-usto.dz.