Journal of Theoretical and Applied Information Technology
31st July 2025. Vol.103. No.14
© Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ELEVATING VOICE DIAGNOSTICS: SAVA UNLEASHES NEW FRONTIERS IN HEALTHY AND PATHOLOGICAL VOICE DETECTION

ABDUL REHMAN ALTAF[1,2*], HAIRULNIZAM MAHDIN[1*], AWAIS MAHMOOD[3], ABDULLAH ALTAF[1], MUHAMMAD HUSSAIN[1], SAJID ISLAM[2]

1Faculty Of Computer Science And Information Technology, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Johor, Malaysia.

2Johns Hopkins Aramco Healthcare (Jhah), Dhahran 34465, Saudi Arabia.

3Faculty Of Applied Computer Science, King Saud University, Riyadh, Saudi Arabia.

E-Mail: hi240034@student.uthm.edu.my, hairuln@uthm.edu.my, mawais@ksu.edu.sa, hi210007@student.uthm.edu.my, gi220008@student.uthm.edu.my, islamsajid@hotmail.com

* Abdul Rehman Altaf and Hairulnizam Mahdin, Are The Corresponding Authors.

ABSTRACT

The detection of pathological voices is a pressing and crucial concern that necessitates a thorough exploration of voice signal properties within healthcare settings. While a plethora of voice features exists, typically leveraged through machine learning techniques to distinguish between healthy and pathological voice signals, yet new voice features are required for more promising results. This study introduces a novel voice feature known as the Sum of the Absolute Values of Amplitudes (SAVA) of voice signals. The development of this feature is meticulously detailed in an algorithmic fashion. To ensure its robustness and reliability, a rigorous evaluation technique, known as K-fold cross-validation, has been employed. This approach not only validates the effectiveness of our feature but also provides insights into its stability and generalization capabilities. Utilizing this feature, a novel framework called the SAVA-Based Classifier (SAVABC) has been devised. In addition, the Guassian Naïve Bayes (GuassianNB) machine learning classifier was chosen for implementation. Extensive voice datasets, comprising both healthy and pathological samples from the Saarbrucken Voice Database (SVD), were utilized for experimentation. The results of the simulations are highly promising, achieving an accuracy of 90.21% for vowel classification and 90.65% for sentence classification. Based on these statistics, we assert that the proposed SAVABC system holds the potential for deployment in real-world healthcare settings to detect pathological voices in patients. This system has the capacity to replace manual assessment methods in healthcare, offering substantial convenience to potential patients.

Keywords: *Healthy Voice, Pathological Voice, Machine Learning, Voice Feature, Signal Processing*

1. INTRODUCTION

Voice is one of the basic instincts of mankind. But unluckily, due to its excessive usage, some people like musicians, politicians, orators, teachers, students and attorneys etc., suffer some problems and disorders[1]. Hence, the exciting domain of speech pathology attracts a phenomenal number of scholars and researchers to investigate the voice disorders. The fact of the matter is that abnormal growth of tissues or masses in vocal folds results in a voice tone that is markedly different from the normal one[2]. The relevant examples of the said irregular growth in vocal folds include cysts, nodules, sulci and polyps[3]. The normal signs of speech pathology are scratchy throat, persistent hoarseness, abnormal volume, and a mitigated capacity for speaking clearly. Figure (1) shows the main components of human voice generation system. Here we will explain how human voice gets generated. Firstly, we inhale air to our lungs. After it, we expel air from lungs by lowering diaphragm located at the bottom of lungs. This action of ours forces air to rush through vocal trachea via epiglottis and causes vocal cords to vibrate. Apart from that, larynx is one of the most delicate parts of human voice generation system. This part comprises of muscles, cartilages, and ligaments. Chief mandate of larynx is to control vocal folds which depends on various voiced and unvoiced conditions. A handsome amount of people suffer from some kind of pathological voice disorder. Around a whopping 25% of world population whose professions necessitate them to speak in a louder tone suffer from these kinds of problems. Apart from that, dysphonia —a typical voice pathology, has spread with such a magnitude that around 7.5 million people in United States alone are suffering from vocal problems[4]. Moreover, approximately 15% patients who visited the King Abdul Aziz University Hospital in Riyadh (Saudi Arabia) were diagnosed with various voice disorders[5]. Additionally, the impact of voice related problems on the teaching professional is even greater than many other professionals. Various surveys in United States have indicated that spread of voice disorders during the whole life of people is 57.7% for teachers and 28.8% for non-teachers[7]. Moreover, in Riyadh (capital of Saudi Arabia), about 33% of female and male teachers suffer from varied voice disorders at some point in their lives[6].

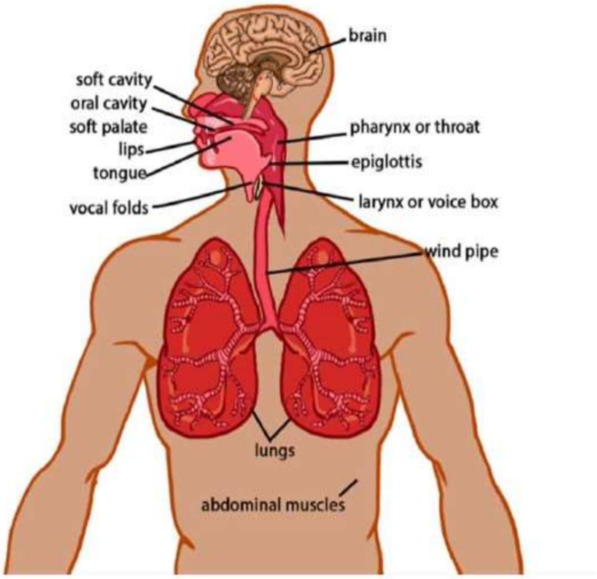


Figure 1. Various Components Of Human Voice Generations System

Voice pathology analysis and its detection have assumed a marked importance in the realm of healthcare. A significant portion of the populace is suffering from this malady including teachers, lawyers, supervisors, actors, singers etc. Apart from that, smoking, air pollution, damage to certain organs and stress are also the causes of this disease[7]. Moreover, a study suggests that around 7.5 million people have voice pathology[8] in USA alone. Two ways exist to detect the voice pathology namely subjective and objective. In the former way, a patient is examined by an expert doctor. Different clinical gadgets are also employed during this process. Apart from that, this process is painful and inconvenient for the patients and doctors alike. Moreover, this procedure also varies from doctor to doctor[9]. One more disadvantage of this method is that it is too much time consuming. As far as the latter method (objective evaluation) is concerned, no particular gadgetry is employed in this method. Rather, signal processing of voice signals of the patients are carried out. In sharp contrast to the first method, there is no inconvenience caused to the patients and doctors. Moreover, this process is speedy. The voice signals are fed to the algorithm. This algorithm gives us the results depending upon the nature of the signals (healthy or pathological)[10]. It is to be noted that although different algorithms are employed but the final decision depends upon the medical doctors.

Researchers have probed various voice features as the literature review indicates. The feature Jitter, for example, uses the successive changes which happen in the periods of the given voice signal[11]. In order to find the Jitter, timing of fundamental period is detected. Through this parameter, voice pathology can be detected as its value is compared with the standard threshold value. This parameter can be used to detect voice pathology by comparing it with a threshold value.

83.2 µs is the threshold value for the detection of pathologies in adults[12,13]. The parameter of Shimmer is also frequently employed for the purpose of voice disability detection[14,15]. It addresses the signal peak values. For the sake of determination of its parameters, it finds starting time of glottal pulses of the given signal and its corresponding magnitude for that sample. Later on, algorithm is subjected for finding the values of all the other parameters. Normalized noise energy (NNE) is an other voice feature which can be defined as a ratio of noise energy and total energy of the given signal[16]. Within the confines of the harmonic, noise energy is actually an average of consecutive minimal value in spectrum. On the other hand, between the harmonics, noise energy is got from given spectrum. In case, due to the effects of Shimmer or Jitter, harmonics get expanded, outside window energy of the harmonics is taken wrongly as noise energy. In this way, noise assessed through NNE got increasing. In order to address this problem, frequency range is changed to get the optimal discrepancy between pathological (glottal cancer) and normal voice.

The mathematics given in[17] is exploited to implement the notion of harmonics-to-noise ratio (HNR). It works on the basis of given voice signal's autocorrelation function. In a sense, HNR is cepstrum-based HNR[18] and it is related to NNE in an inversely fashion. It is actually a quotient of total energy and noise energy. Moreover, Zero Crossing Rate (ZCR)[19] is also used for the classification of healthy and pathological voice signals. The phenomenon of zero crossing occurs if the successive voice samples contain various algebraic signs in the context of discrete-time signals. This is a measure of the different frequency contents of the given voice signal. To put it in other words, it is a measure of frequency of times divided by frame in such a way that its amplitude of the voice signals undergoes the zero value[20]. Linear Frequency Cepstral Coefficient (LFCC) is an other voice feature which is calculated like a filter bank consisting of 40 bands MFCCFB40[21]. Apart from that, the difference lies in the fact that the step of mel frequency warping is ignored[22]. Moreover, in the reported technique, frequency range is realized through a 40 equal-height and equal-width filter banks. Apart from that, bandwidth of all the used filters is kept at 164 Hz. Additionally, these filter banks contain 133 to 6857 Hz as the frequency range. LFCC is calculated through the different steps which are: (1) application of *N*-point Discrete Fourier Transform (DFT) upon the given signal. (2) application of the triangular filtering. (3) computation of filter bank outputs which have already been compressed in a logarithmic fashion. (4) lastly the application of Discrete cosine transform (DCT) over different outputs for getting LFCC FB-40 parameters.

Inspired by the above discussion, this study has ventured to discover a new voice feature based on the amplitudes of the voice signals. The authors of this work observed that the amplitudes of healthy voice signals are greater than their pathological counterparts. By complying with this observation, a new voice feature called sum of absolute values of the amplitudes (SAVA) has been discovered which has the enormous potential to draw a rather clearer line of demarcation between the healthy and pathological voice signals. To ensure the authenticity of this feature, a rigorous evaluation technique, known as *K*-fold cross-validation, has been employed. Apart from that, GuassianNB machine learning classifier[23] has been employed in this study using the voice feature of SAVA. Following is the problem statement of the current study:

Probelm statement:The increasing use of voice-based technologies in areas such as speaker recognition, emotion detection, and health diagnostics demands more effective and efficient voice signal features. While traditional voice features such as Mel Frequency Cepstral Coefficients (MFCC), pitch, and formants have shown considerable success, they may not fully capture all relevant aspects of a voice signal, especially under varying acoustic conditions or in real-time applications. Amplitude-based features of voice signals are often overlooked or underexplored in contemporary research. This raises the need to investigate whether novel, amplitudedriven features can be extracted and used to enhance the performance and robustness of voice processing systems. The core challenge lies in identifying, extracting, and validating a new voice feature derived specifically from the amplitude domain that can outperform or complement existing features in classification, recognition, or analysis tasks.

This study has formulated the following research questions.

RQ1: Can a novel voice feature based on the amplitudes of the signals be discovered?

RQ2: Can the discovered feature be employed to classify the pathological and healthy voice signals?

Having said that, nonetheless, the following points list the contribution of this research work to the body of knowledge in an objective fashion.

1. An acute observation of the amplitudes of the voice signals led us to discover a novel feature of the voice, i.e., sum of absolute values of the amplitudes (SAVA).

2. The discovered feature SAVA in the Step 1 bears an ample promise to draw a relatively better line of demarcation between the healthy and the pathological voice signals.

3. One more salient feature of the modus operandi of the discovery of this new feature is that it is very efficient as far as the time complexity is concerned.

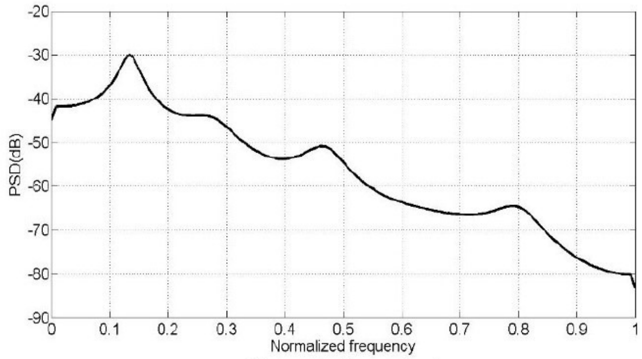
4. Based on this new feature, machine learning algorithm has been employed which rendered very rosy results as far as the validation metrics are concerned.

5. Given the above bullet points, we assert that the proposed framework has the bright prospects for its application in the healthcare settings.

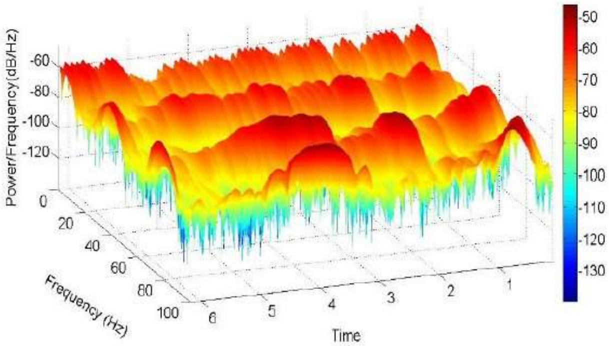
2. RELATED WORK

Voice features characterize the particular voice of a person. These features have the potential for the researchers to decide that the particular voice is healthy or pathological. In the following, a broad treatment would be made for the frequently used voice features by the different researchers.

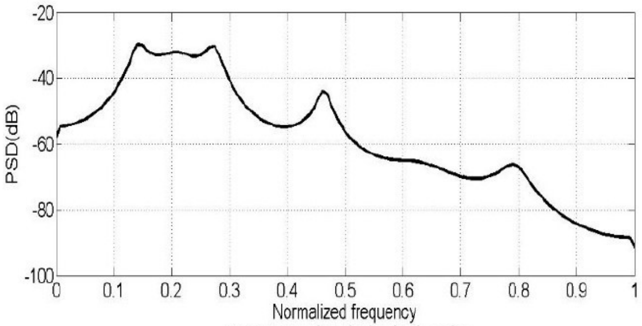
Mel Frequency Cepstral Coefficients (MFCCs) has been universally employed by the researchers for the sake of classification between the healthy and pathological voices[24]. MFCCs is so comprehensive among the other features that it absorbs the shape of the vocal tract in its entirety. In this way, phoneme generated by the vocal tract can be represented with much accuracy. The feature of formant frequency (also called formant analysis) is a yet another significant feature employed by the researchers for the sake of classification. These are resonance frequencies for the vocal tract. Further, these frequencies alter with various configurations of the vocal tract[25]. Formants denote spectral contribution of the given resonance. Apart from that, peaks of given spectrum for the vocal tract response refers to its corresponding formants. Formant plots depict the different peaks at the various frequencies as shown in the Figure (2a). Spectrogram is another frequently used voice feature by the researchers to distinguish between the healthy and the pathological voice signals. In a speech waveform, there is a sequence of different events that change as the time passes by. In this way, this waveform fluctuates wildly during the passage of time. In this way, single Fourier transform is not



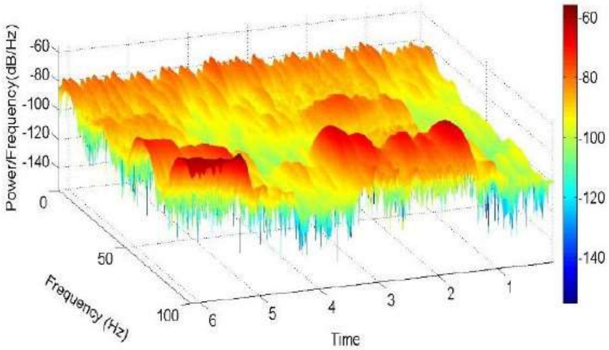
Formants of normal voice



Spectrogram of normal voice sample



Formants of pathological voice



Spectrogram of pathological voice sample

(a) *(B)*

Figure 2. Normal And Pathological Voice Samples: (A) Formants; (B) Spectrograms

potent enough to characterize such kind of fast time varying signal. Therefore, for this purpose, Short-time Fourier transform (STFT) is employed[26]. The STFT comprises of different Fourier transform for different segments of the given waveform in the given sliding window. So, voice signal's spectrogram is obtained through the STFT by. Researchers also depict the spectrograms of the voices in the 3D plots in order to demonstrate the power densities along with the frequency and time (Figure 2b).

Traditionally, the notion of Linear Predictive Coding (LPC) has been exploited for the compression of digital signals for the purpose of transmission and storage. Whereas, this idea has been successfully applied for the analysis of speech signals. Resultantly, it has become very popular as a formant estimator[25]. This method is based on characterizing vocal tract as a linear all-pole infinite impulse response (IIR) filter. Wavelet Analysis is also employed by the researchers for the detection of voice disabilities. The point where it gains advantage over the celebrated Fourier transform is that it renders more precise and accurate information regarding rapid fluctuations of the voice signals in the domain of time. Its modus operandi works like this. It renders the mapping of the time function into two functions, i.e., scale (denoted by *a*) and translation (denoted by *b*)[25]. Apart from that, Hermansky[27] introduced the notion of Perceptual Linear Prediction (PLP). It exploits the psychophysics of the phenomenon of hearing while modeling the human speech. The primary functionality of PLP is to delete the unrelated information present in the speech. Moreover, PLP contains the spectral properties which are translated in such a way that it matches the human auditory system. LPC and PLP employ distinct transfer functions. As an example, LPC model has the assumption about the all-pole transfer function of vocal tract along with particular resonances in the band of analysis. Whereas, PLP's transfer function carries out an approximation of the power distribution with equal magnitudes at all the given frequencies of the band. Rasta Perceptual Linear Prediction (RASTA-PLP) is one more speech feature used frequently by the researchers. In order to calculate this feature, a bandpass termed as

RASTA-PLP is employed. Lower cut-off frequency finds fastest spectral change which is normally overlooked at output. Apart from that, high cut-off frequency finds fast spectral change which is usually preserved in output. Filter's primary job is to abolish frequency that changes rapidly or slowly in the signals of the given voice. The calculation of RASTA-PLP entails some steps. Firstly, critical-band power spectrum is calculated. Then, spectral amplitude is transformed via compressed nonlinear transformation which is static in nature. Afterwards, time trajectory for each transformed spectral component is filtered. In the next step, filtered speech representation is transformed by the expansion of static nonlinear transformation. Then, carry out multiplication by loudness curve and raise it by 0.33 power for simulating power of law for hearing. Lastly, find all-pole model for the spectrum using the classical PLP procedure. Lastly, the work[28] claims to discover a new voice feature based on the maximum peak values and their corresponding lag values from each frame of a voiced signal, but their work is shrouded in mystery as no 'nuts and bolts' in an algorithmic fashion are found in their work.

3.INRODUCTION OF NOVEL VOICE FEATURE BASED ON AMPLITUDES OF VOICE SIGNALS

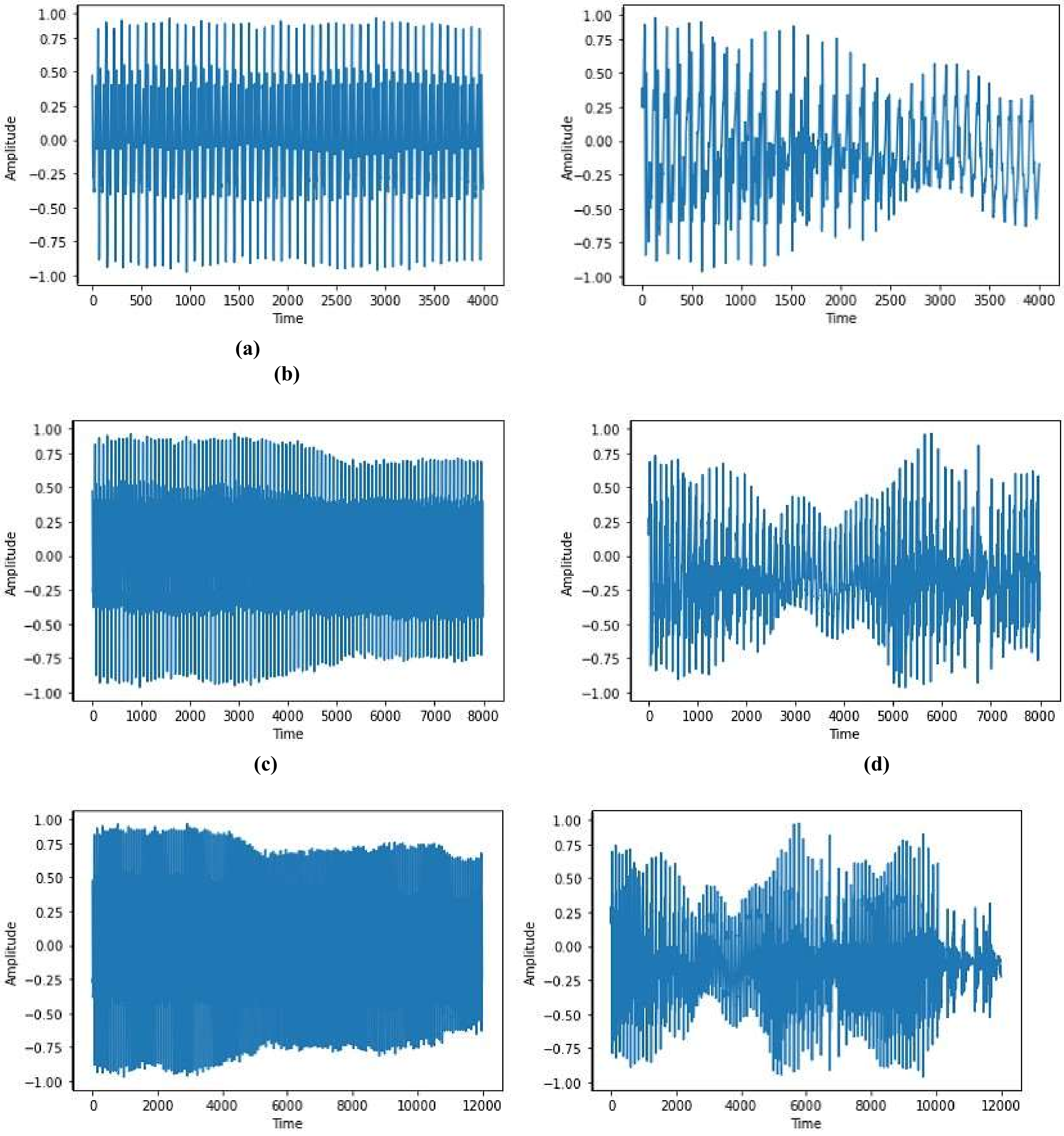
3.1 Motivation of the Study

An acute observation regarding the amplitudes of the healthy and pathological voices led us to discover a novel feature of the voice which almost clearly draws a line of demarcation between these two kinds of voices. One can see from the Figure (3) that the amplitudes of healthy voices have greater heights as compared to their pathological counterparts. Apart from that, Tables (1) and (2) show the sum of the absolute values of the amplitudes of healthy and the pathological voices for selected sixty files in the database. The grand average values for the metric SAVA are 9894 and 2973 for the healthy and pathological voices respectively. Moreover, the cumulative average (CA) of the average values of 9894 and 2973 calculates to be 6433. In the Tables (3) and (4), a comparison has been made for the metric SAVA against this cumulative average value CA. Apart from that, the frequencies of the correction have been calculated. According to the Table (3), the frequency of correction is 46. Its corresponding percentage value comes out to be 46/60 ×100% ≈ 76.67%. In the same fashion, Table (4) shows that the frequency of correction is 58. In percent form, it is 58/60 ×100% ≈ 96.67%. By aggregating both of these averages, the percentage comes out to be 104/120 ×100% ≈ 86.67%. This percentage 86.67% is reasonable enough to dig the problem further and implement it for the detection of pathological voice signals.

3.2 SAVA Unveiled: Pioneering an Algorithm for Introducing a Novel Voice Feature

Algorithm (1) finds the average value of the amplitudes of training data. It is to be noted that this training data Algorithm (1) finds the average value of the amplitudes of training data. It is to be noted that this training data consists of healthy and pathological voices with m and n files. Line 1 initializes the variable sum_o f _amplitudes to zero. Line 2 iterates the f or loop for m+n times. On the line 3, load function of the librosa module of the Python language has been invoked by passing the ith value of the array. This function returns an array signal of amplitudes and sampling rate sampling_rate for the selected file. On the line 4, the function librosa.util.normalize normalizes the audio signal and returns the array normalized_signal. Normalization is a common audio processing technique used to scale the amplitudes of the audio samples to a specific range, between -1 and 1. Line 5 once again iterates the f or loop for len (normalized_signal) times. Line 6 checks whether the value normalized_signal[i] is negative? If it is, it is being made positive on the line 7. Line 10 evaluates the sum of all the amplitudes of the given data and assigns the result to the variable temp. All these calculated results are being accumulated in the variable sum_o f _amplitudes on the line 11. Lastly, SAVA has been calculated on the line 13.

The Algorithm (2) takes both the training_data and training_data along with the number of files. The pairs (m,n) and (p,q) refer to numbers of the healthy and pathological voice files for training_data and training_data



(a)

(b)

(c)

(d)

(e)

(f)

Figure 3. *Amplitude Distribution Of Healthy And Pathological Voices Along With The Values Of SAVA: (A) Healthy Voice For First 4000 Units, 1291.90; (B) Pathological Voice For First 4000 Units, 894.65; (C) Healthy Voice For First 8000 Units, 2474.44; (D) Pathological Voice For First 8000 Units, 1744.50; (E) Healthy Voice For First 12000 Units, 3583.70; (F) Pathological Voice For First 12000 Units, 2425.70*

Table 1. Sum Of Absolute Values Of Amplitudes Of The Healthy Voices

File Number	**SAVA**	**File Number**	**SAVA**	**File Number**	**SAVA**	**File Number**	**SAVA**
1	12373	16	7777	31	5183	46	6339
2	13278	17	10870	32	6897	47	6645
3	8220	18	11194	33	14429	48	12884
4	9529	19	6512	34	15291	49	6975
5	6916	20	8638	35	16584	50	11130
6	5619	21	12067	36	12444	51	3922
7	7858	22	6040	37	5598	52	20049
8	12277	23	7232	38	10327	53	13260
9	7420	24	7623	39	3904	54	5736
10	12909	25	4493	40	7124	55	9463
11	13514	26	7938	41	7718	56	13542
12	21426	27	4548	42	7790	57	14182
13	13141	28	4976	43	14174	58	8258
14	22055	29	19277	44	8541	59	5155
15	5063	30	13597	45	6095	60	9649
Average of SAVA	11439		8852		9473		9812
Grand Average of SAVA							9894

Table 2. Sum Of Absolute Values Of Amplitudes Of The Pathological Voices

File Number	SAVA	File Number	SAVA	File Number	SAVA	File Number	SAVA
1	3977	16	3293	31	2177	46	2973
2	6946	17	2102	32	2031	47	2379
3	1066	18	2387	33	3225	48	2241
4	1224	19	1502	34	3352	49	4219
5	4218	20	3168	35	2298	50	2929
6	2317	21	2398	36	2649	51	2069
7	3656	22	1443	37	1702	52	6838
8	1319	23	4797	38	3375	53	4119
9	1723	24	2945	39	2865	54	1987
10	3542	25	4412	40	2063	55	6248
11	2907	26	2069	41	2625	56	5238
12	2776	27	2412	42	1092	57	2635
13	1588	28	2083	43	2407	58	4453
14	2476	29	3113	44	2559	59	2249
15	4436	30	4585	45	4636	60	1915
Average of SAVA	2944		2847		2603		3499
Grand Average of SAVA							2973

respectively. The algorithm (2) labels the given voice signals based on the Cumulative Average (CA) value of the given voice file. Line 1 invokes the Algorithm AverageValueCal with the parameters of training_data,m and n. Majority of steps of Algorithm (2) is the same as that of

Algorithm (1). Lines (11-15) decide that whether any ith file of testing_data is pathological or not? If the i f condition on the line 11 evaluates to be true (means the value of CA is less than that of SAVA), then the Boolean value of true is getting assigned to the variable pathological and vice versa.

3.3 K-Fold Cross Validation of the Proposed Voice Feature

To ensure the robustness and reliability of the voice feature SAVA, we employ here a rigorous evaluation technique known as K-fold cross-validation29 —a widely recognized method for appraising the performance of machine learning models and features. This approach will not only validate the efficacy of the proposed feature but will also provide insights into its stability and generalization capabilities. This technique divides the given dataset into K folds (or subsets) and performs training and testing iteratively on these folds, thus allowing us to evaluate our feature's performance under various condi tions. The following steps characterize the particular modus operandi of this technique.

- *Data Collection*: A comprehensive dataset was collected comprising of voice recordings from individuals with and without pathological voice conditions. This dataset, no doubt, served as the foundation for our evaluation.
- *Data Preprocessing*: Before proceeding with feature extraction, the collected voice data underwent preprocessing to remove noise, handle missing values, and standardize the format.
- *Feature Engineering*: The newly developed voice feature is based on the amplitudes of the voice signals. The unique feature SAVA was extracted given the pathological and healthy voice signals.
- *Split Dataset into Training and Testing*: To facilitate cross-validation, the dataset was divided into a training set and a testing set. This separation ensured that the model was trained on one portion of the data and evaluated on another.
- *K-Fold Splitting*: We adopted a *K*-fold cross-validation approach, with typically *K* = 10. This strategy enhances model evaluation by dividing the dataset into *K* equal-sized subsets or folds.
- Use (K-1) Folds for Training: During each iteration, (K-1) out of the K subsets (folds) are used for training the model. This process is repeated K times, with each fold taking a turn as the testing set. It ensures that the model is exposed to diverse training data and provides a robust assessment of its performance on various subsets, aiding in better generalization evaluation.

Algorithm 1:AverageValueCal

```
Input: training_data, m, n
Output: SAVA
1: sum_o f _amplitudes ← 0
2: for i ← 1 to m+n do
3:     [signal,sampling_rate] ← librosa.load(training_data[i])
4:     normalized_signal = librosa.util.normalize(signal)
5:     for i ← 1 to len(normalized_signal) do
6:         if normalized_signal[i] < 0 then
7:             normalized_signal[i] ← −normalized_signal[i]
8:         end if
9:     end for
10:    temp ← sum(normalized_signal)
11:    sum_o f _amplitudes ← sum_o f _amplitudes+temp
12: end for
13: SAVA ← sum_o f _amplitudes/(m+n)
```

Table 3. Comparison of SAVA with CA for the healthy voices

File Number	SAVA	SAVA > CA?	File Number	SAVA	SAVA >CA?	File Number	SAVA	SAVA >CA?	File Number	SAVA	SAVA > CA?
1	12373	✓	16	7777	✓	31	5183	×	46	6339	×
2	13278	✓	17	10870	✓	32	6897	✓	47	6645	✓
3	8220	✓	18	11194	✓	33	14429	✓	48	12884	✓
4	9529	✓	19	6512	✓	34	15291	✓	49	6975	✓
5	6916	✓	20	8638	✓	35	16584	✓	50	11130	✓
6	5619	×	21	12067	✓	36	12444	✓	51	3922	×
7	7858	✓	22	6040	×	37	5598	×	52	20049	✓
8	12277	✓	23	7232	✓	38	10327	✓	53	13260	✓
9	7420	✓	24	7623	✓	39	3904	×	54	5736	×
10	12909	✓	25	4493	×	40	7124	✓	55	9463	✓
11	13514	✓	26	7938	✓	41	7718	✓	56	13542	✓
12	21426	✓	27	4548	×	42	7790	✓	57	14182	✓
13	13141	✓	28	4976	×	43	14174	✓	58	8258	✓
14	22055	✓	29	19277	✓	44	8541	✓	59	5155	×
15	5063	×	30	13597	✓	45	6095	×	60	9649	✓
Fre.of Correction		**13**			**11**			**11**			**11**
Grand Fre. of Correction											**46**

Table 4. Comparison Of SAVA With CA For The Pathological Voices

File Number	SAVA	SAVA < CA?	File Number	SAVA	SAVA < CA?	File Number	SAVA	SAVA < CA?	File Number	SAVA	SAVA < CA?
1	3977	✓	16	3293	✓	31	2177	✓	46	2973	✓
2	6946	×	17	2102	✓	32	2031	✓	47	2379	✓
3	1066	✓	18	2387	✓	33	3225	✓	48	2241	✓
4	1224	✓	19	1502	✓	34	3352	✓	49	4219	✓
5	4218	✓	20	3168	✓	35	2298	✓	50	2929	✓
6	2317	✓	21	2398	✓	36	2649	✓	51	2069	✓

7	3656	✓	22	1443	✓	37	1702	✓	52	6838	×
8	1319	✓	23	4797	✓	38	3375	✓	53	4119	✓
9	1723	✓	24	2945	✓	39	2865	✓	54	1987	✓
10	3542	✓	25	4412	✓	40	2063	✓	55	6248	✓
11	2907	✓	26	2069	✓	41	2625	✓	56	5238	✓
12	2776	✓	27	2412	✓	42	1092	✓	57	2635	✓
13	1588	✓	28	2083	✓	43	2407	✓	58	4453	✓
14	2476	✓	29	3113	✓	44	2559	✓	59	2249	✓
15	4436	✓	30	4585	✓	45	4636	✓	60	1915	✓
Fre. of Correction		**14**			**15**			**15**			**14**
Grand Fre. of Correction											**58**

Algorithm 2: Labelling voice signal as pathological or healthy based on SAVA

```
Input: training_data,testing_data, m, n, p, q
Output: pathological
SAVA ← AverageValueCal(training_data,m,n)
for i ← 1 to p+q do
    [signal,sampling_rate] ← librosa.load(testing_data[i])
    normalized_signal = librosa.util.normalize(signal)
    for i ← 1 to len(normalized_signal) do
        if normalized_signal[i] < 0 then
            normalized_signal[i] ← −normalized_signal[i]
        end if
    end for
    CA ← sum(normalized_signal)
    if CA < SAVA then
        pathological ← true
    else
        pathological ← false
    end if
end for
```

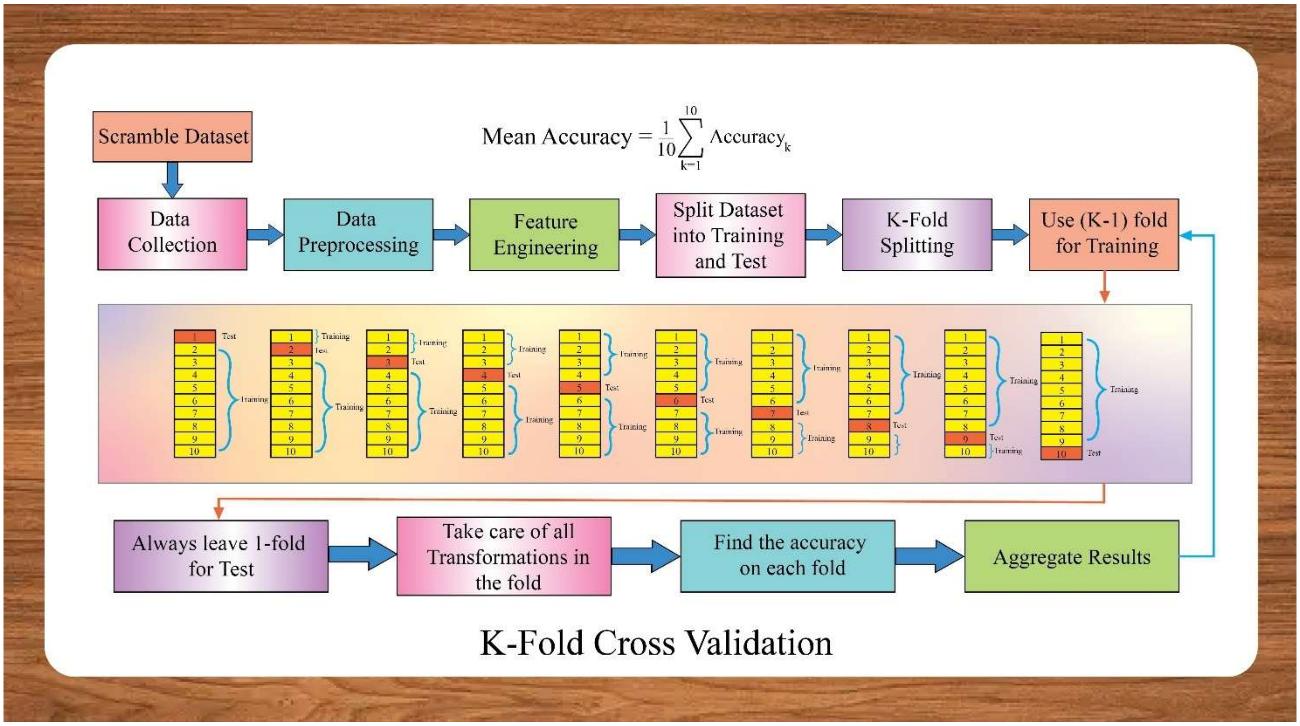
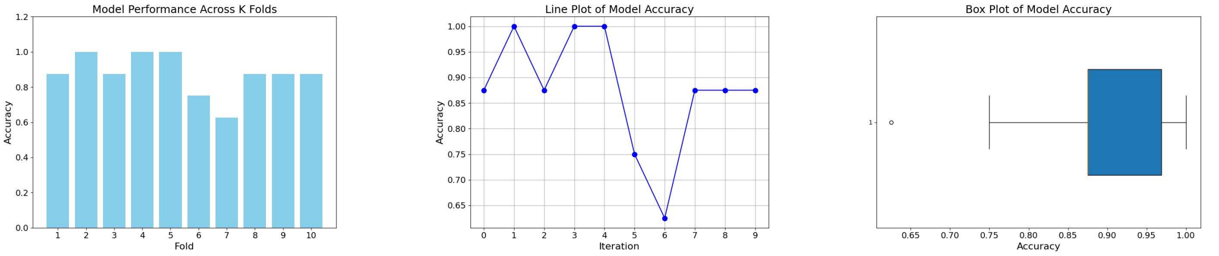


Figure 4. K-Fold Cross Validation

The Figure (4) describes all the steps of this technique in detail. We got the scores of cross validations as {0.875, 1.0, 0.875, 1.0, 1.0, 0.75, 0.625, 0.875, 0.875, 0.875} with the mean of 0.875 and standard deviation of 0.1118. This aptly signals towards the intrinsic robustness and reliability of the newly discovered voice feature across different subsets of the dataset. Besides, the Figure (5) draws the results using the instruments of Bar plot, Line plot and Box plot.



(a) **(b)** **(c)**

Figure 5. Performance Of The SAVA Through Different Graphs: (A) Bar Plot; (B) Line Plot; (C) Box Plot

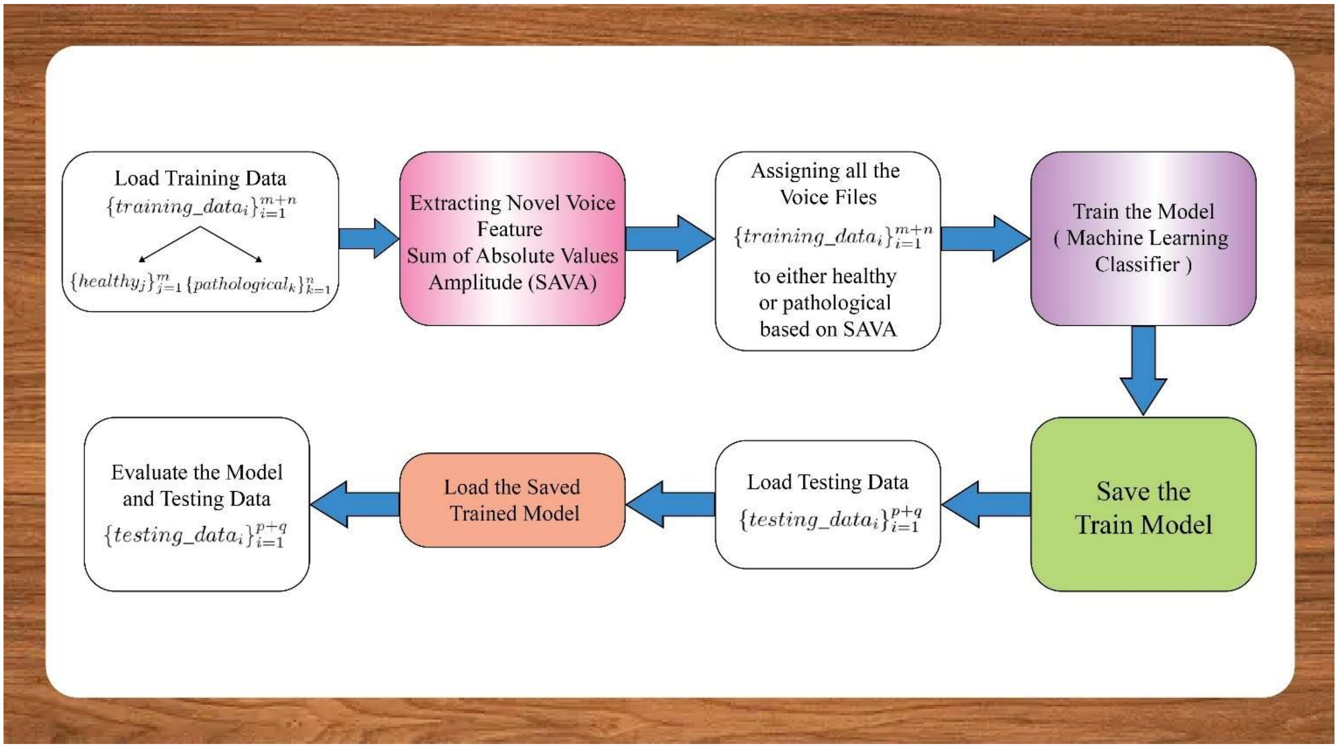


Figure 6. Proposed Methodology

4. PROPOSED METHODOLOGY FOR THE CLASSIFICATION BETWEEN HEALTHY AND PATHOLOGICAL VOICES

In this section, we will discuss the proposed methodology for the classification of the healthy and pathological voice signals using the novel voice feature SAVA. Figure (6) draws a flowchart of the proposed framework for classifying between the healthy and pathological voice signals. The work comprises of some stages. In the Stage 1, traing data $\{training\_data_i\}_{i=1}^{m+n}$ is getting loaded. According to the diagram, it consists of *m* files $\{healthy_j\}_{j=1}^{m}$ for the healthy voices and *n* files $\{pathological_k\}_{k=1}^{n}$ for the pathological voices. These two sets of files are fed to the two Algorithms (1) and (2) in order to extract the novel voice feature of SAVA (Stage 2). This is the very feature which will do the seminal work for the discrimination between the given voice signals. In the Stage 3, all the files $\{training\_data_i\}_{i=1}^{m+n}$ are getting labelled as healthy or pathological based on the feature of SAVA. In the next Stage 4, model of the machine learning classifier, i.e., GuassianNB has been trained. Stage 5 is saving the trained model. In the Stage 6, the testing data $\{testing\_data_i\}_{i=1}^{p+q}$ is being loaded. Now the saved trained model is being loaded in the Stage 7. In the last Stage 8, the model is being evaluated based on the testing data.

4.1 The Voice Database

To demonstrate the practical utility and effectiveness of the proposed framework, we have chosen SVD (Saarbrucken Voice Database) for the sake of experimentation[30]. This database has a rich collection of the voice recordings of about 2000 persons. Besides, it contains the 687 (259 males and 428 females) recordings of healthy voices and 1354 (627 males and 727 females) recordings of pathological voices. Besides, in these recordings, 71 different pathologies have been catered. Additionally, these voice recordings have been developed with a resolution of 16-bit and a frequency of 50 kHz. Apart from that, the duration of these samples spans from 1 to 3 seconds. Moreover, the average age of the speakers is 15 years. SVD database comprises of two kinds of recordings. In the first kind, recordings of vowels /a/, /u/, and /i/ have been carried out. These recordings have been uttered with distinct intonations. For instance, the vowel "a" has been uttered in varied ways like /a/_neutral (a_n), /a/_high (a_h), and /a/_low (a_l). Same holds true for other vowel letters as well. As far as the second type of recording is concerned, it denotes a continuous recording carried out in the German language. For instance, "Guten Morgen, wie geht es Ihnen?" It means "Good morning, how are you?" Furthermore, we allocated 80% of voice samples for training and the remaining 20% for testing process.

4.2 Simulation Results and Analyses

The required experiments were conducted with the tool of Python 3. Moreover, GaussianNB algorithm has been employed for the sake of classification between the pathological and healthy voice signals as described earlier. Apart from that, the evaluation measures used are accuracy, precision, recall (sensitivity), F-measures, G-mean, specificity, and execution time as shown in equation (1) to equation (5)[31–33]. In the following, we will describe the varied measures.

1. TP (True Positive): given voice signal is of pathological nature and algorithm reports it as of pathological nature.
2. TN (True Negative): given voice signal is of healthy nature and algorithm reports it as of healthy nature.
3. FP (False Positive): given voice signal is of healthy nature but algorithm reports it as of pathological nature.
4. FN (False Negative): given voice signal is of pathological nature but algorithm reports it as of healthy nature.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Sensitivity(Recall) = \frac{TP}{TP+FN} \tag{3}$$

$$F-Measure = \frac{2\times Precision\times Recall}{Recall+Precision} \tag{4}$$

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

The experimentation process demonstrated that both the novel voice feature SAVA and GaussianNB algorithm are potent enough to clearly differentiate between the healthy voices and the pathological voices. Moreover, the above assertion holds true regarding the varied vowels alongwith distinct pitches and sentences from SVD database. Table (5) gives the various results of the applied GaussianNB algorithm for the SVD database. 90.21% and 82.87% came out to be the accuracy and G-mean for /a_h/ database. Apart from that, 92% precision (for /a_l/) and 90% recall (for all /a/) has been achieved. It to be noted that all /a/ denotes the combination of a_h, a_l, and a_n. Additionally, 86% F-measure (for all /i/) and 94.67% specificity (for /a_l/) has been obtained. These results have been visually demonstrated in the Figure 7. Moreover, 1.1 seconds came out to be the training and testing time for the GaussianNB classifier for /u_n/ database. Moreover, 81.07% came out to be the least accuracy for all vowels and the sentences in the SVD database. This explains the fact that GaussainNB algorithm experiences hurdles for classifying the voice signals with different sentences and vowels. An array of evaluation measures exists but accuracy is much universal in its character and orientation. Hence, this study has chosen this universal metric to judge the performance of the proposed framework to differentiate between the healthy and pathological voice signals. Comparative analysis has been carried out based on sustained vowel letters like /a/, /i/, and /u/ and other running sentences. It is to be noted that the vowels are produced in various pitches like high, low, and normal. The suggested method has been compared with the work in[34] for three vowels as described earlier along with three pitches. Moreover, the comparison has been carried out with the work[35] for /a_h/ pitch. Besides, suggested algorithm has been compared with the method given in[36] for /a_h/ and /u_h/ pitches. Likewise, the suggested algorithm has been compared with the works of[37] and[38]. The works[37] and[39] used /a_n/ pitch only; and study[38] employed all vowels and the sentences uttered in the normal pitches. A comparison based on different metrics has been made in the Table (6) between suggested algorithm and other published works in all sessions of SVD database.

According to the results, suggested work beats the selected methods in terms of accuracy for the detection of pathological voice signals. Moreover, the results obtained have been depicted visually in the Figure (7).

Researchers of voice pathology detection system normally employ the metrics like accuracy, specificity, and sensitivity only. In contrast to that, the current work has also used computational time. Moreover, our work has been compared with the work[40] in terms of the execution time. Besides, the algorithm[40] employed the CNN algorithm in the pathology detection for the voice signals using vowel /a/ produced in the normal pitch. Execution time for the CNN was 2.54 seconds for testing voice signal. In contrast to that, execution time for suggested algorithm is 1.1 seconds. It beats CNN by 56.29%.

Table (7) shows the comparison of execution time between SAVABC and CNN based on the vowel /a_n/.

Table 5. Results Of The Proposed SAVABC Classifier For The SVD Database

Accuracy	Precision	Recall	F-measure	G-mean	Specificity	Time (Sec)
			a_h			
90.21%	88%	77%	83%	82.87%	93.07%	1.1
			a_l			
89.81%	92%	75%	86%	83.32%	94.67%	1.2
			a_n			
91.20%	79%	82%	77%	82.29%	92.08%	1.1
			All_a			
82.89%	79%	90%	83%	83.32%	74.67%	1.2
Accuracy	Precision	Recall	F-measure	G-mean	Specificity	Time (Sec)
			i_h			
83.43%	85%	73%	77%	77.71%	93.76%	1.2
			i_l			
88.52%	84%	72%	76%	78.87%	92.81%	1.1
			i_n			
87.23%	83%	82%	71%	87.23%	91.34%	1.2
			All_i			
83.89%	89%	91%	86%	83.32%	74.67%	1.2
Accuracy	Precision	Recall	F-measure	G-mean	Specificity	Time (Sec)
			u_h			
82.24%	81%	74%	79%	80.90%	88.12%	1.1
			u_l			
81.07%	83%	89%	88%	84.12%	89.78%	1.1
			u_n			
83.67%	87%	88%	87%	81.89%	90.72%	1.1

All_u						
84.90%	85%	88%	82.89%	80.25%	91.89%	1.2
Accuracy	Precision	Recall	F-measure	G-mean	Specificity	Time (Sec)
Sentences						
90.65%	88%	90%	87%	84.26%	90.83%	1.1

Table 6. Comparison Based On Different Metrics Between SAVABC And Other Methods

Methods	**Accuracy**	**Methods**	**Accuracy**	**Methods**	**Accuracy**	**Methods**	**Accuracy**
a_h		**a_n**		**a_l**		**All_a**	
SAVABC	91.20%	SAVABC	90.21%				
		GMM[34]	67% 68.08%	SAVABC	89.81%	SAVABC	82.89%
DNN37							
GMM	66.6%	SVM[41]	85.77%				
SVM[35]	71.12%	DNN[39]	82.01%	GMM[34]	65.6%	GMM[34]	71.8%
CNN[36]	73%	SVM[38]	74.32%				
SVM[42]	84.37%	Ref.[43]	81.33%				
Methods	**Accuracy**	**Methods**	**Accuracy**	**Methods**	**Accuracy**	**Methods**	**Accuracy**
i_h		**i_n**		**i_l**		**All_i**	
		SAVABC	87.23%				
SAVABC	83.43%	GMM$_{34}$	64.5%	SAVABC	88.52%	SAVABC	83.89%
GMM[34]	64%	SVM[38]	72.29%	GMM[34]	64.2%	GMM[34]	71%
Methods	**Accuracy**	**Methods**	**Accuracy**	**Methods**	**Accuracy**	**Methods**	**Accuracy**
u_h		**u_n**		**u_l**		**All_u**	
SAVABC	82.24%	SAVABC	83.67%				
GMM$_{34}$	64%	GMM$_{34}$	63.4%	SAVABC	81.07%	SAVABC	84.90%
CNN[36]	63%	SVM[38]	71.45%	GMM[34]	64.6%	GMM[34]	71.5%

Methods	**Accuracy**
Sentences	
SAVABC	90.65%
SVM[38]	76.19%
DNN[44]	87.4%
GMM[45]	80.2%

Table 7. Execution Time Of The Proposed SAVABC And CNN

Methods	**Computational time (Sec)**
SAVABC	1.1
CNN[40]	2.54

5.DISCUSSION

Traditionally, varied invasive surgical treatments are carried out to treat the voice problems. During the painful, expensive and time consuming procedures of laryngoscopy, laryngeal electromyography, and stroboscopy, medical doctors insert probes of different types into the mouth of the potential patients. Diametrically opposite to this subjective procedure, there exists an objective procedure in which the signal processing of the voice is carried out. Here, the voice features enjoy the pivotal position The reason of this is that these features are pregnant of the fact that whether the voice is healthy or pathological. Plethora of voice features have already been discovered. Apart from that, those features are frequently being employed by the machine learning experts for the sake of classification between the healthy and pathological voices. This study has ventured to discover an other voice feature.

The underlying traits of this voice feature include 1) It is simply the sum of the absolute values of the amplitudes of feature is very efficient 3) The classifier gave very promising and competitive results. At the tale of the Introduction section, we formulated two research questions. Here we will analyze those research questions in the light of the findings of this research. The first question was about the discovery of the novel voice feature based on the amplitudes of the voice signals. As we plotted the graphs of the amplitudes of both the healthy and pathological voices, we figured out a pattern in their attitude. The amplitudes of the healthy voices had longer heights and depths as compared to their pathological counterparts. To avoid the inherent danger of getting cancelled the positive and negative heights, we took their absolute values so that the potential voice feature becomes more pronounced. We named the new feature as SAVA meaning sum of absolute values of the amplitudes. As far as the second research question is concerned the given voice signal 2) The computational time of this.

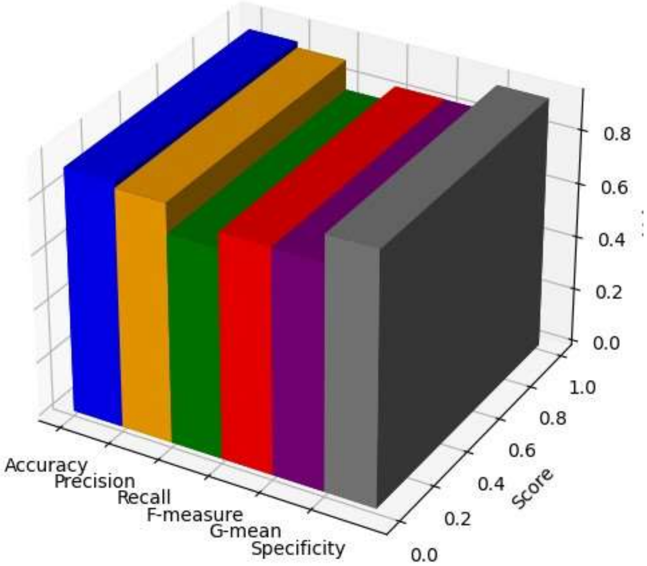


Figure 7. Visual Demonstration Of Validation Metrics

, this study has formulated a framework by incorporating the newly discovered feature of the voice. The name of the framework is SAVABC meaning SAVA based classifier. Clearly, the proposed framework gave very competitive results. Moreover, the study made comparison with the other published works. We assert that the suggested work is better than the studies in some aspects, we took from the literature for the sake of comparison.

6. CONCLUSION AND FUTURE DIRECTIONS

The employment of signal processing for the detection of pathological voice is a hot research issue these days. An array of voice features exist which carry out a seminal task to differentiate the pathological voices from the healthy ones. This work has discovered a novel voice feature (SAVA) based on the sum of the absolute values of the amplitudes of the given voice signals. To demonstrate the robustness and potency of the reported feature, this study has suggested a novel framework SAVABC (SAVA Based Classifier). In the core of the reported framework, GuassianNB machine learning classifier and the newly discovered voice feature SAVA have been embedded. Apart from that, the voice dataset has been taken from the SVD database. Moreover, the vowels and the running German sentences have been used in the experiments. All the validation metrics like accuracy, specificity, G-mean rendered very promising results. Given these stats, we assert that the proposed framework can be used in some real world scenario like hospitals, clinics and other healthcare settings.

As far as the future research directions are concerned, many other classifiers can be tested to get the better results than ours. Moreover, the newly discovered voice feature SAVA can be used in combination of the already discovered voice features to improve the validation metrics. Apart from that, other voice databases can also be employed for the better results. Additionally, other voice maladies like Cyst, Polyy and Paralysis can also be investigated using this new voice feature.

REFERENCES

[1]. Islam, R., Tarique, M. & Abdel-Raheem, E. A survey on signal processing based pathological voice detection techniques. *IEEE Access* 8, 66749–66776 (2020).

[2]. Behrman, A. Common practices of voice therapists in the evaluation of patients. *J. Voice* 19, 454–469 (2005).

[3]. Muhammad, G. *et al.* Formant analysis in dysphonic patients and automatic arabic digit speech recognition. *Biomed. engineering online* 10, 1–12 (2011).

[4]. Al-Nasheri, A., Muhammad, G., Alsulaiman, M. & Ali, Z. Investigation of voice pathology detection and classification on different frequency regions using correlation functions. *J. Voice* 31, 3–15 (2017).

[5]. Ali, Z., Hossain, M. S., Muhammad, G. & Sangaiah, A. K. An intelligent healthcare system for detection and classification to discriminate vocal fold disorders. *Futur. Gener. Comput. Syst.* 85, 19–28 (2018).

[6]. Malki, K. H. *et al.* Voice disorders among saudi teachers in riyadh city. *Saudi J. Otorhinolaryngol. Head Neck Surg.* 12, 31 (2010).

[7]. Syed, S. A., Rashid, M. & Hussain, S. Meta-analysis of voice disorders databases and applied machine learning techniques. *Math. Biosci. Eng.* 17, 7958–7979 (2020).

[8]. Al-Dhief, F. T. *et al.* Voice pathology detection and classification by adopting online sequential extreme learning machine. *IEEE Access* 9, 77293–77306 (2021).

[9]. Song, P. Assessment of vocal cord function and voice disorders. *Princ. practice interventional pulmonology* 137–149 (2013).

[10]. Thomas, F. *et al.* Discrete time speech signal processing (2001).

[11]. Ma, Z. *et al.* Towards the objective speech assessment of smoking status based on voice features: a review of the literature. *J. Voice* (2021).

[12]. Teixeira, J. P., Ferreira, D. & Carneiro, S. M. Análise acústica vocal-determinação do jitter e shimmer para diagnóstico de patalogias da fala. In *6º Congresso LusoMoçambicano de Engenharia, 3º Congresso de Engenharia de Moçambique*, 6º (INEGI, 2011).

[13]. Boersma, P. *et al.* Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the institute of phonetic sciences*, vol. 17, 97–110 (Amsterdam, 1993).

[14]. Vikram, C. & Umarani, K. Pathological voice analysis to detect neurological disorders using mfcc and svm. *Int. J. Adv. Electr. Electron. Eng* 2, 87–91 (2013).

[15]. Srinivasan, V., Ramalingam, V. & Arulmozhi, P. Artificial neural network based pathological voice classification using mfcc features. *Int. J. Sci. Environ. Technol.* 3, 291–302 (2014).

[16]. Kasuya, H., Ogawa, S. & Kikuchi, Y. An adaptive comb filtering method as applied to acoustic analyses of pathological voice. In *ICASSP'86. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 11, 669–672 (IEEE, 1986).

[17]. Boersma, P. *et al.* Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the institute of phonetic sciences*, vol. 17, 97–110 (Amsterdam, 1993).

[18]. Krom, G. d. A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *J. Speech, Lang. Hear. Res.* 36, 254–266 (1993).

[19]. Syed, S. *et al.* Inter classifier comparison to detect voice pathologies. *Math. Biosci. Eng.* 18, 2258–2273 (2021).

[20]. Fonseca, E. S. *et al.* Acoustic investigation of speech pathologies based on the discriminative paraconsistent machine (dpm). *Biomed. Signal Process. Control.* 55, 101615 (2020).

[21]. Lee, J.-Y. Experimental evaluation of deep learning methods for an intelligent pathological voice detection system using the saarbruecken voice database. *Appl. Sci.* 11, 7149 (2021).

[22]. Shrawankar, U. & Thakare, V. M. Techniques for feature extraction in speech recognition system: A comparative study. *arXiv preprint arXiv:1305.1145* (2013).

[23]. Pushpakumar, R. *et al.* A novel approach to identify dynamic deficiency in cell using gaussian nb classifier. In *2022 7th International*

Conference on Communication and Electronics Systems (ICCES), 31–37 (IEEE, 2022).

[24]. Tirronen, S., Kadiri, S. R. & Alku, P. The effect of the mfcc frame length in automatic voice pathology detection. *J. Voice* (2022).

[25]. Islam, R., Tarique, M. & Abdel-Raheem, E. A survey on signal processing based pathological voice detection techniques. *IEEE Access* 8, 66749–66776 (2020).

[26]. Alan, V. & Schafer, R. Fourier transform and fourier analysis of signals using the discrete fourier transform. In *Discrete-Time Signal Processing*, 855–859 (Pearson, 2009).

[27]. Hermansky, H. Perceptual linear predictive (plp) analysis of speech. *J. Acoust. Soc. Am.* 87, 1738–1752 (1990).

[28]. Al-Nasheri, A. *et al.* Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions. *Ieee Access* 6, 6961–6974 (2017).

[29]. Pal, K. & Patel, B. V. Data classification with k-fold cross validation and holdout accuracy estimation methods with 5 different machine learning techniques. In *2020 fourth international conference on computing methodologies and communication (ICCMC)*, 83–87 (IEEE, 2020).

[30]. Reid, J., Parmar, P., Lund, T., Aalto, D. K. & Jeffery, C. C. Development of a machine-learning based voice disorder screening tool. *Am. J. Otolaryngol.* 43, 103327 (2022).

[31]. Albadr, M. A. A., Tiun, S., Al-Dhief, F. T. & Sammour, M. A. Spoken language identification based on the enhanced self-adjusting extreme learning machine approach. *PloS one* 13, e0194770 (2018).

[32]. Albadr, M. A. A. *et al.* Optimised genetic algorithmextreme learning machine approach for automatic covid19 detection. *PloS one* 15, e0242899 (2020).

[33]. Albadr, M. A. A. *et al.* Optimised genetic algorithmextreme learning machine approach for automatic covid19 detection. *PloS one* 15, e0242899 (2020).

[34]. Martínez, D., Lleida, E., Ortega, A., Miguel, A. & Villalba, J. Voice pathology detection on the saarbrücken voice database with calibration and fusion of scores using multifocal toolkit. In *Advances in Speech and Language*

[35]. *Technologies for Iberian Languages: IberSPEECH 2012 Conference, Madrid, Spain, November 21-23, 2012. Proceedings*, 99–109 (Springer, 2012).

[36]. Mohammed, M. A. *et al.* Voice pathology detection and classification using convolutional neural network model. *Appl. Sci.* 10, 3723 (2020).

[37]. Rueda, A. & Krishnan, S. Augmenting dysphonia voice using fourier-based synchrosqueezing transform for a cnn classifier. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6415–6419 (IEEE, 2019).

[38]. Harar, P. *et al.* Voice pathology detection using deep learning: a preliminary study. In *2017 international conference and workshop on bioinspired intelligence (IWOBI)*, 1–4 (IEEE, 2017).

[39]. Kadiri, S. R. & Alku, P. Analysis and detection of pathological voice using glottal source features. *IEEE J. Sel. Top. Signal Process.* 14, 367–379 (2019).

[40]. Al-Dhief, F. T. *et al.* Voice pathology detection and classification by adopting online sequential extreme learning machine. *IEEE Access* 9, 77293–77306 (2021).

[41]. Mohammed, M. A. *et al.* Voice pathology detection and classification using convolutional neural network model. *Appl. Sci.* 10, 3723 (2020).

[42]. Verde, L., De Pietro, G. & Sannino, G. Voice disorder identification by using machine learning techniques. *IEEE access* 6, 16246–16255 (2018).

[43]. Albadr, M. A. A. *et al.* Fast learning network algorithm for voice pathology detection and classification. *Multimed. Tools Appl.* 1–32 (2024).

[44]. Hegde, K. J., Shenoy, K. M. & Devaraja, K. A novel stacked model for classification of vocal cord paralysis over imbalanced vocal data. *IEEE Access* (2025).

[45]. Moon, J. & Kim, S. An approach on a combination of higher-order statistics and higher-order differential energy operator for detecting pathological voice with machine learning. In *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, 46–51 (IEEE, 2018).

[46]. Ali, Z. *et al.* Intra-and inter-database study for arabic, english, and german databases: do conventional speech features detect voice pathology? *J. Voice* 31, 386–e1 (2017).