A Novel Voice Feature AVA and its Application to the Pathological Voice Detection Through Machine Learning

Abdulrehman Altaf[1]*, Hairulnizam Mahdin[2]*, Ruhaila Maskat[3]*,
Shazlyn Milleana Shaharudin[4], Abdullah Altaf[5], Awais Mahmood[6]
Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn
Malaysia, Batu Pahat, Johor, Malaysia[1,2]
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA
Shah Alam, Selangor, Malaysia[3]
Faculty of Science and Mathematics, Universiti Pendidikan,
Sultan Idris, Tanjong Malim, 35900, Perak, Malaysia[4]
Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn, Malaysia, Batu Pahat, Johor, Malaysia[5]
College of Applied Computer Science, King Saud University, Riyadh, Saudi Arabia[6]

***Abstract*—Voice pathology is a universal problem which must be addressed. Traditionally, this malady is treated by using the surgical instruments in the varied healthcare settings. In the current era, machine learning experts have paid an increasing attention towards the solution of this problem by exploiting the signal processing of the voice. For this purpose, numerous voice features have been capitalized to classify the healthy and pathological voice signals. In particular, Mel-Frequency Cepstral Coefficients (MFCC) is a widely used feature in speech and audio signal processing. It denotes spectral characteristics of a voice signal, particularly of human speech. The modus operandi of MFCC is too time-consuming, which goes against the hasty and urgent nature of the modern times. This study has developed a yet another voice feature by utilizing the average value of the amplitudes (AVA) of the voice signals. Moreover, Gaussian Naive Bayes classifier has been employed to classify the given voice signals as healthy or pathological. Apart from that, the dataset has been acquired from the SVD (Saarbrucken Voice Database) to demonstrate the workability of the proposed voice feature and its usage in the classifier. The machine experimentation rendered very promising results. Particularly, Recall, F1 and accuracy scores obtained, are 100%, 83% and 80%, respectively. These results vividly imply that the proposed classifier can be installed in various healthcare settings.**

Keywords*—Pathological voice; healthy voice; voice feature; amplitudes; machine learning

I. Introduction

People whose professions cause them to speak louder than normal, often suffer from some kind of voice pathology. These people may include lawyers, auctioneers, motivational speakers, legislators, singers, teachers, etc. This pathology, in turn, leads to tiredness, infections of voice tissue, face soreness, muscular dystrophy and others [1]. Apart from that, this pathology casts a negative impact upon the voice functionality and vibration regularity which sometimes leads to the increment in the vocal noise. Normal voices turned to be weak, tense, and hoarse which influences quality of voice [2]. Traditionally, voice pathology detection methods are tendentious in their character and orientation. They are based on subjective matters [3]. For instance, in the different hospital settings, an auditory-perceptual assessment is employed which includes visual laryngostroboscopy assessment [4]. In this painful process of diagnosis, a series of clinical examinations are employed for the auditory-perceptual parameters to appraise the severity of the voice malady [5]. These appraisals are subjective and are very sensitive to the sensitivity of the parameters involved. Moreover, they happen to be very much time consuming which is not in line with the current standards of quality [6]. One more disadvantage of this method is that the patients have to be present in the hospital physically which is, of course, not feasible for the patients with critical conditions.

In sharp contrast to that, there exists an objective evaluation of the voice pathology using signal processing of the voice. In particular, the signals of patients' voice are processed to conclude whether the patient concerned is suffering from the pathology or not? No surgical treatment is employed in this method. Moreover, this procedure can also work upon the inaudible sounds [1]. These methods do not depend upon the human decisions. A patient needs not be physically present in the healthcare centres since only his/her voice is required to reach to the decision. So, the voice recording can also be shared through the internet. Upon surveying the relevant literature, one will find that different voice pathology databases exist for the sake of objective evaluation of voice pathologies. Among these, the most common include Arabic Voice Pathology Database (AVPD) [7], Saarbruecken Voice Database (SVD) [8] and the Massachusetts Eye and Ear Infirmary Database (MEEI) [9]. Upon surveying the literature about the pathological voice detection, researchers have used varied voice features and diverse machine learning classifiers for discriminating between the pathological and healthy voice signals [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21]. The work [10] wrote a robust voice pathology detection algorithm using the theory of Deep Learning. In order to maximize the accuracy of classification, the pre-trainend Convolutional Neural Network

(CNN) has been employed over the dataset of voice pathology. Besides, this work used SVD as a dataset for testing their work. The accuracy claimed by the reported study is 95.41%. Moreover, F1-Score and Recall scores were calculated to be 94.22% and 96.13%, respectively. In an other study [11], the voice pathology detection system was developed in the realm of smart healthcare. In this particular work, the voice data was taken through the IoT gadgetry, i.e., electroglottography (EGG) and microphones for capturing the EGG and voice signals. The voice feature spectrogram was employed in this particular study. These spectrograms were got from the reported signals and were given as an input to the pre-trained CNN. Moreover, the features obtained through the usage of CNN were mixed and later on processed through short long-term memory network (bi-directional in nature). The accuracy claimed by the said study was 95.65%. In an another research [12], the authors employed an Online Sequential Extreme Learning Machine (OSELM) as the classification algorithm in their work for detection of pathological voice signals. An other prominent feature of this study is the employment of long sentences instead of the single vowel letters for the sake of discrimination between the pathological and healthy voice signals. Three types of voice pathologies were addressed namely cyst, polyp, and paralysis. The accuracy achieved was 91.17%. Apart from that, precision and recall scores were 94% and 91%, respectively. The work is reported to give a high capability for detecting the pathological voice signals in the real-time clinical settings.

Many voice features have been discovered by the academicians and other researchers. Some of these include formant frequency [22]. Formant frequencies are a sort of resonance frequencies. These frequencies change with various vocal tract configurations [23]. Commonly, these formants denote the spectral contribution of the given resonances. Apart from that, peaks of these spectra about the local tract responses refer to the corresponding formants. The various plots of these formants depict the different peaks at the various frequencies. Spectrogram of a voice signals [24] is yet another voice feature. They are a kind of a waveform comprising of various events which change as the time goes by. Owing to the fact that they vary with the time, hence they fluctuate and exhibit the spectral properties. This is the reason that a single Fourier transform [25], [26] is humble to capture such kind of speedy time varying signals. Hence, for this purpose, a short-time Fourier transform (STFT) was used. STFT comprises of different Fourier transform for the pieces of the given waveform. The feature of linear predictive coding (LPC) is also used by machine learning experts to differentiate between the pathological and healthy voice signals [27], [28]. Initially, LPC was designed for compressing the digital signals for the efficient storage and transmission of the digital data. In current times, this feature is frequently being employed to draw a line of discrimination between the healthy and pathological voice signals. Moreover, this method models vocal tract in the form of linear all-pole infinite impulse response (IIR) filter.

Calculation of these features is mathematically intensive. Apart from that, they consume a lot of precious processing time which is not in line with the demands of the current era. So, we require simple but powerful voice features to do the job.

In this work, a novel voice feature by observing the behavior of amplitudes of the voice signals has been discovered. In particular, the average value of the amplitudes of the voice signal has been determined. This average value is potent enough to differentiate between the healthy and pathological voice signals. Moreover, this feature has been embedded in the machine learning algorithm to classify the two kinds of signals. The machine experimentation rendered very competitive results. Moreover, these results are better than many of those published in the literature.

Having said that, the following salient features characterize the contribution of this work to the exciting field of voice signal processing and machine learning:

- A novel voice feature based on the average value of amplitudes of the given voice signals has been determined.
- The voice feature found in the above bullet has been exploited in the machine learning algorithm to draw a rather clearer line of demarcation between the pathological and healthy voice signals.
- The proposed method rendered very competitive results. Moreover, it beats many results of the published works.

Rest of the paper has been fashioned like this. In Section II, the particular modus operandi employed in MFCC feature extraction has been explained. Section III describes the proposed methodology. In particular, the way novel voice feature has been determined, has been explained in detail. Afterwards, the reported feature has been embedded in the proposed framework to differentiate between the healthy and pathological voice signals. In Section IV, the results have been described and compared with the other published researches in the literature. Section V closes the paper with the concluding remarks and other possible research directions.

II. Related Work

Traditionally, MFCC has been employed by the machine learning experts to draw a line of demarcation between the healthy and pathological voice signals. MFCC is actually a feature selection method which plays a very critical role to distinguish the pathological voice signals from their healthy counterparts. Normally, three kinds of features are there for the recognition of the sound patterns. They are time domain, frequency domain and time-frequency domain [29]. Cepstral domain features are retrieved after taking their fast Fourier transform (FFT) of the amplitude's logarithm from spectrum data [30]. Since MFCCs closely resemble human auditory system, so their inherent power is normally harnessed for the speech recognition in the diverse problems [31]. MFCCs are normally got through power spectrum of sound signals with the short-term windowing after taking cosine transform of logarithmic power spectrum over Mel filter banks [32]. A standard modus operandi for extracting the MFCCs from the given audio signals have been depicted in the Fig. 1. Firstly, windowing functions, like Hanning and Hamming windows, are normally employed through some degree of overlapping for capturing local spectral characteristics. Secondly, various signals from the different frames are subjected to operation of

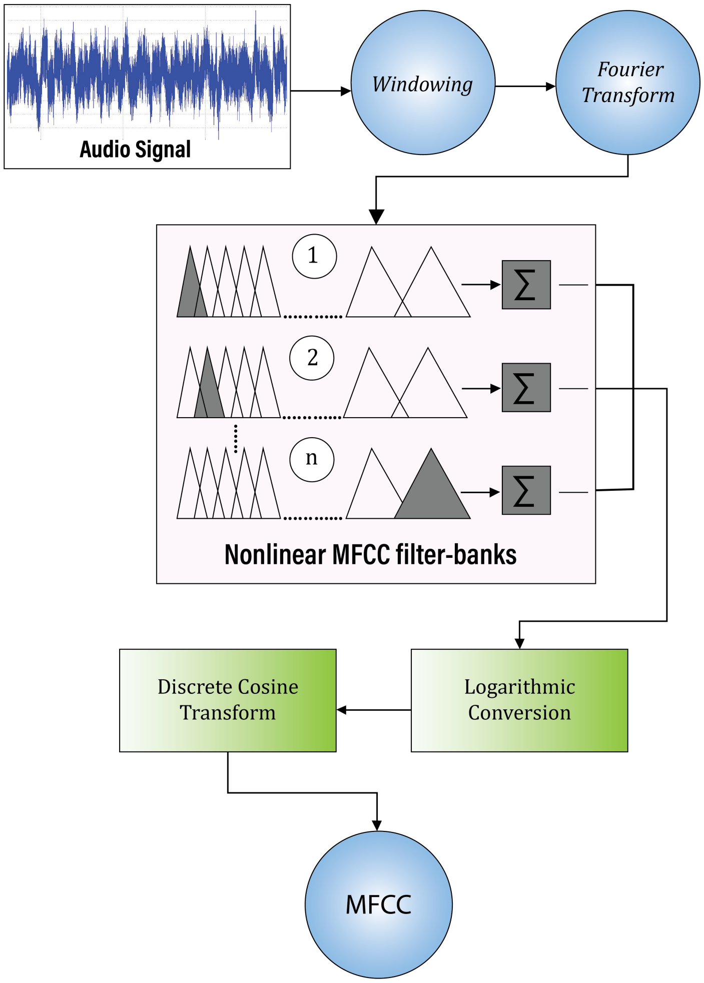


Fig. 1. MFCC feature extraction mechanism. given audio signals have been bifurcated into the overlapping frames along with some fixed intervals and weights have been given through hamming window.

discrete Fourier transform (DFT) for the sake of extracting information from the frequency domain. Thirdly, frequency domain data are filtered through a number of bandpass filters which are normally equal to designed number of the MFCC resolution (64,128,256,512). It is to be noted that centre frequencies of bandpass filters are spaced in uniformity on Mel scale $M(f)$ [33].

$$M(f) = \frac{1000 \quad ln(1+\frac{f}{700})}{ln(1+\frac{1000}{700})} \approx 1127 \quad ln(a+\frac{f}{700}) \qquad (1)$$

In this equation, f denotes frequency term and $M(f)$ denotes Mel scale. Moreover, this equation converts boundaries of filter bank to Mel scale. As soon as centre frequencies are distributed in a uniform fashion on Mel scale, values are converted back to frequency domain which renders the triangular filters. After that energies $MF(t)$ of corresponding filter banks are computed by taking sum of energies in bandpass filters. Finally, MFCC coefficients are found through the application of discrete cosine transform (DCT) to the filtered energies from the triangular bandpass filters [34].

$$MFCC_{i,j} = \frac{1}{T}\sum_{k=1}^{T} log[MF(k)]cos[\frac{2\pi}{T}(k+\frac{1}{2})j] \qquad (2)$$

In this equation, $MFCC_{i,j}$ refers to the j^{th} $MFCC$ coefficient of i^{th} frame. Apart from that, $1 \leq i \leq N$ and $1 \leq j \leq M$. They represent the indices of $MFCC$. Moreover, $MF(k)$ is Mel filter bank amplitude of k^{th} filter. Apart from that, Table I sheds light on the varied studies carried out. In this table, one can examine the different studies based on the number of samples taken, phonemes, pathological condition of the patients, the classifier employed, the feature used and lastly the findings. Here will describe few studies in more details. In study [35], normal and pathological samples taken were 60 and 402, respectively. Besides, the vowels were taken as phonemes to apply the classifier. Apart from that, the voices of the patients were suffering from the pathological conditions of structure lesions and neoplasm. Additionally, the classifiers selected for this particular study were Support Vector Machine (SVM), Gaussian Mixture Modelling (GMM) and Deep Neural Network (DNN). The feature upon which the distinction was made between the healthy and pathological voice was MFCC. As far as the findings and outcomes of this study were concerned, SVM outperformed GMM. Besides, the classifier DNN rendered the highest accuracy. The study [36] took 56 normal samples of voices and 67 pathological samples. Moreover, the phonemes employed in this particular study were '/ah/'. The pathological conditions of the patients concerned were that they were suffering from the Parkinson's disease, Vocal cord paralysis and cerebral demyelination. The classifier and the feature selected were SVM and MFCC. The accuracy obtained in this study was 93%. The last row of this table describes these parameters for the proposed study. It is to be noted that, we employed AVA as a voice feature for the sake of classification between the healthy and the pathological voice signals.

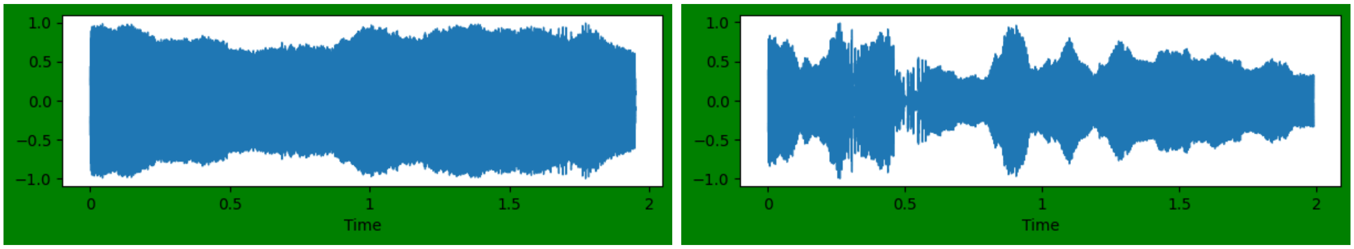
One can note that this traditional method of extracting the voice feature is very complicated and mathematically intensive. We require simple but powerful voice features to draw a clearer line of demarcation between the given voice signals.

III. Proposed Methodology

In this work, a novel voice feature consisting of average value of the amplitudes (AVA) of the given voice signal has been determined. This voice feature has been, in turn, employed to distinguish between the given healthy and pathological voice signals. Fig. 2 draws the amplitudes of the healthy and pathological voice signals. Fig. 2a and 2b refer to the signals for the healthy and pathological voices. One can clearly observe that amplitudes in the positive and negative sides for the healthy signals are greater than their pathological counterparts. This is the very observation through which the novel voice feature has been determined. In particular, the average value of the amplitudes (both positive and negative) of the healthy and pathological voice signals has been found. Their average values have been further averaged. This average has been used while training our model. Once the model gets trained, testing phase have been employed.

TABLE I. OVERVIEW TABLE

Sr. #	Study	Samples	Phonemes	Pathological condition	Classifier	Feature	Findings
1	Ref. [35]	Normal: 60 Pathological: 402	Vowels	Structural lesions, neoplasm	SVM, GMM, DNN	MFCC	SVM outperforms GMM DNN provides the highest accuracy
2	Ref. [36]	Normal: 56 Pathological: 67	Vowel '/ah/'	Parkinson's disease, Vocal cord paralysis cerebral demyelination	SVM	MFCC	Highest accuracy of 93%
3	Ref. [37]	Pathological: 60	Japanese vowel	Breathiness, Roughness, asthma and strain	Higher-Order Local Autocorrelation (HLAC)	Auto Regressive, (AR)-HMM, Feed Forward Neural Networks (FFNN)	87.75% accuracy
4	Ref. [38]	Normal: 53 Pathological: 602	Vowel '/ah/', Rainbow passage (German, ,Japanese and English)	Hyper function, Paralysis, Anterior-poster squeezing, Gastric reflux	PRAAT	Pitch, Jitter Shimmer and HNR	Efficient for English, not efficient for German and Japanese
5	Ref. [39]	Pathology: 65 Normal: 13	Spanish vowel	Dysphonia, Hyernasality and Dysarthria	Hidden Markov Model (HMM)	Nonlinear parameter, entropy	99% accuracy
6	Ref. [40]	Normal: 49 Pathological: 87	Vowel '/a/'	Dysphonia	Pitch Detection Algorithm(PDA)	Pitch	Better than PRAAT
7	Proposed	Normal: 50 Pathological: 50	Vowels	Dysphonia	GaussianNB	AVA	Accuracy is 80%



(a)

(b)

Fig. 2. Amplitudes of healthy and pathological voices: (a) Healthy voice; (b) Pathological voice.

A. Voice Feature Based on Average Value of the Amplitudes (AVA)

This subsection finds the average value of the amplitudes of the given voice signals. The flowchart for extracting the novel voice feature AVA has been depicted in the Fig. 3. Call the Algorithm 1 with the parameters of training data (*TD*), number of healthy voice files (*P*) and the number of pathological voice files (*Q*). Algorithm 1 works like this. The *for* loop of line 1 iterates for $P + Q$ times. In each iteration, it loads the i^{th} file of the training data TD by using the load function of the Python module librosa. It returns the stream of amplitudes $data$ and the sampling rate $sampling\_rate$. Line 3 initializes a variable $sum\_of\_amplitudes$ to zero. Lines 4 to 6 take the absolute value of the amplitudes of the voice signal carried by the array $data$. Lines 7 and 8 find the average value of the amplitudes and assigns this value to the variable $temp$. Line 9 accumulates the average values in the variable $sum\_of\_amplitudes$. Lastly, line 10 finds the grand average value of all the average values of the amplitudes of the given training data and assigns this value to the variable $AvgValue$. Algorithm 2 has been designed to train the data based on the average value found through the Algorithm 1. Line 1 invokes the Algorithm $AvgValue$ with the parameters TD, P and Q and assigns the result to the variable AV. Lines 7 and 8 find the average value of the amplitudes of the given i^{th} file. If this value is less than the AV (line 9), this particular file is being labelled as pathological (line 10), otherwise, it is being labelled as healthy (line 12).

Algorithm 1: *AvgValue* Calculation of Average Value of the Amplitudes of Voice Signals

```
   Input: $TD$, $P$, $Q$
   Output: $AvgValue$
1  for $i \leftarrow 1$ to $P + Q$ do
2      $[data, sampling\_rate] \leftarrow librosa.load(TD[i])$
3      $sum\_of\_amplitudes \leftarrow 0$
4      for $j \leftarrow 1$ to $len(data)$ do
5          if $data[j] < 0$ then
6              $data[j] \leftarrow -data[j]$
7      $temp \leftarrow sum(data)$
8      $temp \leftarrow \frac{temp}{len(data)}$
9      $sum\_of\_amplitudes \leftarrow$
         $sum\_of\_amplitudes + temp$
10 $AvgValue \leftarrow \frac{sum\_of\_amplitudes}{P+Q}$
```

B. Healthy and Pathological Voice Classifier Based on the Average Value of the Amplitudes

Proposed pathological voice classifier has been presented in the Fig. 4. On the left half of the figure, training procedure has been illustrated in a step by step fashion. The process gets sparked with the given training data which comprises of both healthy and pathological voice files. In the next stage, the amplitudes have been extracted from the given voice signals. It is followed by the calculation of the average value of the amplitudes (both in the positive and negative directions). Based

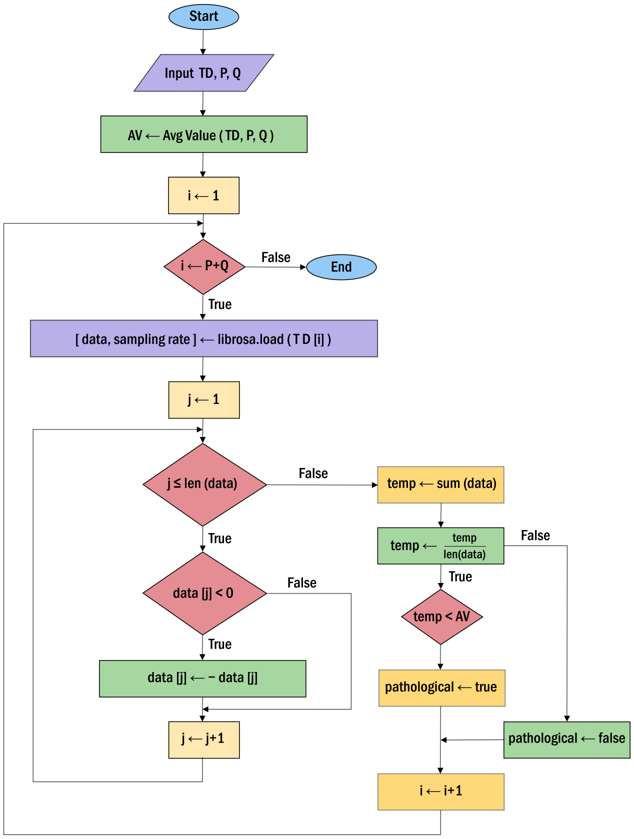


Fig. 3. Flowchart of extracting AVA.

Algorithm 2: Labelling voice signal as pathological or healthy based on $AvgValue$.

```
Input: $TD$, $P$, $Q$
Output: $pathological$
$AV \leftarrow AvgValue(TD, P, Q)$
for $i \leftarrow 1$ to $P + Q$ do
    $[data, sampling\_rate] \leftarrow librosa.load(TD[i])$
    for $j \leftarrow 1$ to $len(data)$ do
        if $data[j] < 0$ then
            $data[j] \leftarrow -data[j]$
    $temp \leftarrow sum(data)$
    $temp \leftarrow \frac{temp}{len(data)}$
    if $temp < AV$ then
        $pathological \leftarrow true$
    else
        $pathological \leftarrow false$
```

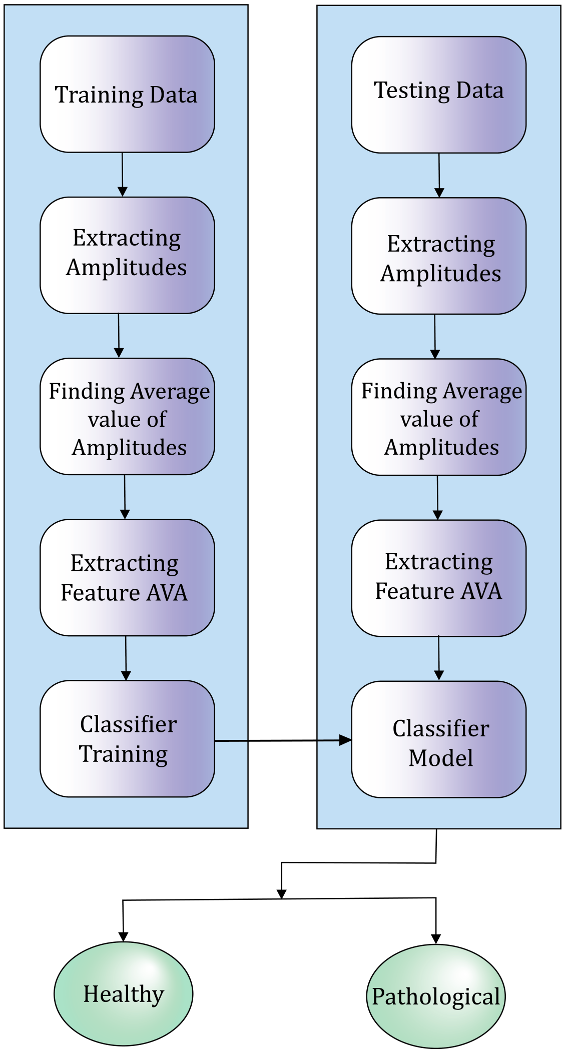


Fig. 4. AVA based pathological voice classifier.

on this average value AVA, the GaussianNB classifier has been trained. Same process has been repeated on the right half of the figure which is pertinent to the testing data. Lastly, the classifier model outputs whether the particular voice signal is healthy or of pathological character.

IV. Simulation Results

The proposed framework was simulated on the Python 3 software. We have taken 80% voice files as a training data and 20% voice files as a testing data. Additionally, these files have been taken from the SVD database. The proposed research project has used this database for the sake of experimentation. It has a collection of voice recordings of around 2000 people. To put it specifically, it contains 687 healthy voice comprising of 259 males and 428 females. Moreover, there are the recordings of 1354 pathological voices comprising of 627 males and 727 females. It is to be noted that all these recordings contain 71 different pathologies. Moreover, these recordings were sampled at 50 kHz frequency along with a 16-bit resolution. Additionally, average age of the speakers is

TABLE II. Results (in percent form) of Different Validation Metrics using the Proposed Methodology and other Methods

Method	Feature	Accuracy	Precision	Recall	F1 Score	Specificity	G-mean
Ref. [10]	-	95.41	-	96.13	94.22	-	-
Ref. [11]	-	93.94	95.08	94.87	94.93	-	-
Ref. [12]	MFCC	91.17	94.0	91.0	87.0	97.67	87.55
Ref. [13]	Peak and Lag	88.70	-	88.69		88.71	-
Ref. [13]	Entropy	82.01	-	73.90		89.72	-
Ref. [14]	MPEG-7	99.994	-	73.90		89.72	-
Ref. [15]	LLE+CD	90.0	-	88.0	-	98.0	-
Ref. [16]	MDVP	76.0	-	45.0	-	93.0	-
Proposed	AVA	80.0	71.0	100.0	83	60.0	100.0

around 15 years. Apart from that, 1 to 3 seconds is the duration of these voice samples.

As far as the machine learning algorithm is concerned, GaussianNB algorithm was chosen to classify the healthy and pathological voice signals. Moreover, in this study, we have chosen these validation measures: accuracy, precision, recall (sensitivity), F-measures, G-mean, and specificity as shown in Eq. 3 to Eq. 7 [41], [42]. The following describes the various measures which are frequently employed in the literature.

1) FP (False Positive): The voice signal under consideration is of healthy character but algorithm declares it as of pathological character.
2) FN (False Negative): The voice signal under consideration is of pathological character but algorithm declares it as of healthy character.
3) TP (True Positive): The voice signal is of pathological character and algorithm declares it as of pathological character.
4) TN (True Negative): The voice signal is of healthy character and algorithm declares it as of healthy character.

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

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Sensitivty(Recall) = \frac{TP}{TP + FN} \tag{5}$$

$$F - Measure = \frac{2 \times Precision \times Recall}{Recall + Precision} \tag{6}$$

$$Specificity = \frac{TN}{TN + FP} \tag{7}$$

The Table II shows the results of the proposed study. Apart from that, this table also draws a comparison between the suggested work and other published researches found in the literature. As can be seen from the table, we got both the Recall and G-mean scores as 100% which validates and confirms the very idea of AVA, we conceived before launching this project. Apart from that, we got an accuracy of 80% which beats some of the published works in the literature. Moreover, one can see that various machine learning experts have used the voice feature of MFCC, Peak and Lag, Entropy, MPEG-7, LLE+CD and MDVP. Unfortunately, our study could only beat the study [16] as far as the metric of accuracy is concerned. However, our results of Recall, G-Mean and F1 score are very competitive.

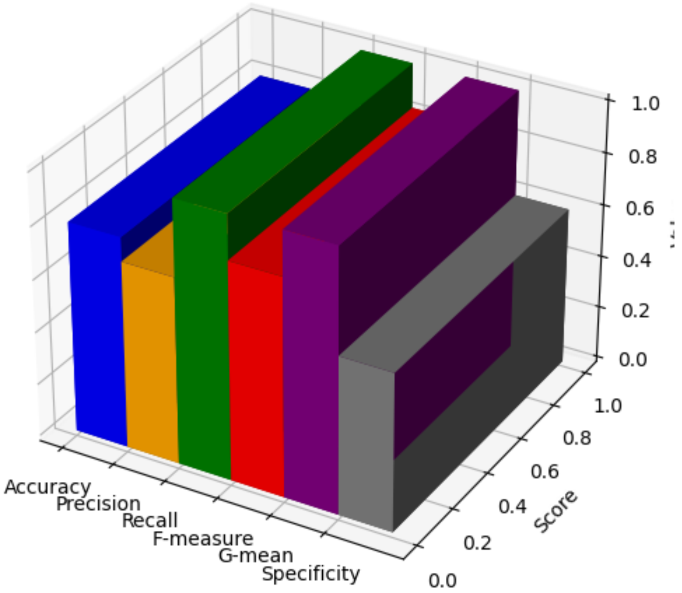


Fig. 5. Visual demonstration of validation metrics.

Besides, the Fig. 5 depicts the results of the proposed study in a graphical form which is more intuitive.

V. Discussion

Many voice features have been investigated as the literature over the pathological and healthy voice classifiers suggests. These features are input to the machine learning classifiers in order to draw a line of separation between the given pathological and healthy voice files of the different patients. The authors of this study observed a pattern after drawing and putting side by side the graphs of the pathological and healthy voices. The amplitudes of the healthy voice signals went higher as compared to their pathological counterparts. This was the very point which was further investigated. In this way, a novel voice feature termed as Average Values of the Amplitudes was found which was further imported to the machine learning classifier in order to draw a clear line of demarcation between the healthy and pathological voice signals. This study obtained Recall and G-mean scores as 100% while the accuracy achieved reached to the tune of 80%. This outcomes vividly imply that the suggested voice feature is potent enough to predict the healthy and pathological voice signals. Moreover, we contend that the proposed voice feature can be extracted with faster speed as compared to the other features like MFCC. MFCC has twelve parameters whereas the proposed one has only one parameter. As far as the limitations of the proposed framework are concerned, it can't differentiate all the voice pathologies.

VI. Conclusion

By observing an underlying pattern in the given voice signals, a novel voice feature AVA has been developed in this study based on the varying values of the amplitudes of the healthy and pathological voice signals. Although a plethora of voice features already exist when one peruses the literature but this newly developed voice feature is very simple and robust to draw a clearer line of demarcation between the healthy and the pathological voice signals. This feature has been exploited while using the machine learning algorithm to detect the pathological voices. The simulation and the machine experimentation rendered very promising results. In particular, we got the Recall and G-mean score as 100% while the accuracy achieved by this study is 80%. We assert that the proposed voice classifier can be installed in some real world healthcare setting to reap its intrinsic benefits. As a future work, other machine learning classifiers and the voice databases would be investigated to examine the workability of the novel voice feature AVA, this study has produced.

Acknowledgment

I would like to express my sincere gratitude to my supervisor and the reviewers for their invaluable support and feedback. Without their guidance and input, this paper would not have been possible. Additionally, I would like to thank my family and friends for their unwavering encouragement throughout this research journey.

References

[1] F. T. Al-Dhief, N. M. A. Latiff, N. N. N. A. Malik, N. S. Salim, M. M. Baki, M. A. A. Albadr, and M. A. Mohammed, "A survey of voice pathology surveillance systems based on internet of things and machine learning algorithms," *IEEE Access*, vol. 8, pp. 64 514–64 533, 2020.

[2] M. A. Mohammed, K. H. Abdulkareem, S. A. Mostafa, M. Khanapi Abd Ghani, M. S. Maashi, B. Garcia-Zapirain, I. Oleagordia, H. Alhakami, and F. T. Al-Dhief, "Voice pathology detection and classification using convolutional neural network model," *Applied Sciences*, vol. 10, no. 11, p. 3723, 2020.

[3] J. Hillenbrand, R. A. Cleveland, and R. L. Erickson, "Acoustic correlates of breathy vocal quality," *Journal of Speech, Language, and Hearing Research*, vol. 37, no. 4, pp. 769–778, 1994.

[4] N. Saenz-Lechon, J. I. Godino-Llorente, V. Osma-Ruiz, and P. Gómez-Vilda, "Methodological issues in the development of automatic systems for voice pathology detection," *Biomedical Signal Processing and Control*, vol. 1, no. 2, pp. 120–128, 2006.

[5] M. Markaki and Y. Stylianou, "Using modulation spectra for voice pathology detection and classification," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2009, pp. 2514–2517.

[6] M. S. Hossain, G. Muhammad, and A. Alamri, "Smart healthcare monitoring: a voice pathology detection paradigm for smart cities," *Multimedia Systems*, vol. 25, pp. 565–575, 2019.

[7] N. Q. Abdulmajeed, B. Al-Khateeb, and M. A. Mohammed, "A review on voice pathology: Taxonomy, diagnosis, medical procedures and detection techniques, open challenges, limitations, and recommendations for future directions," *Journal of Intelligent Systems*, vol. 31, no. 1, pp. 855–875, 2022.

[8] J.-N. Lee and J.-Y. Lee, "An efficient smote-based deep learning model for voice pathology detection," *Applied Sciences*, vol. 13, no. 6, p. 3571, 2023.

[9] N. Q. Abdulmajeed, B. Al-Khateeb, and M. A. Mohammed, "A review on voice pathology: Taxonomy, diagnosis, medical procedures and detection techniques, open challenges, limitations, and recommendations for future directions," *Journal of Intelligent Systems*, vol. 31, no. 1, pp. 855–875, 2022.

[10] M. A. Mohammed, K. H. Abdulkareem, S. A. Mostafa, M. Khanapi Abd Ghani, M. S. Maashi, B. Garcia-Zapirain, I. Oleagordia, H. Alhakami, and F. T. Al-Dhief, "Voice pathology detection and classification using convolutional neural network model," *Applied Sciences*, vol. 10, no. 11, p. 3723, 2020.

[11] 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. 89 198–89 209, 2021.

[12] F. T. Al-Dhief, M. M. Baki, N. M. A. Latiff, N. N. N. A. Malik, N. S. Salim, M. A. A. Albader, N. M. Mahyuddin, and M. A. Mohammed, "Voice pathology detection and classification by adopting online sequential extreme learning machine," *IEEE Access*, vol. 9, pp. 77 293–77 306, 2021.

[13] A. Al-Nasheri, G. Muhammad, M. Alsulaiman, Z. Ali, K. H. Malki, T. A. Mesallam, and M. F. Ibrahim, "Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions," *Ieee Access*, vol. 6, pp. 6961–6974, 2017.

[14] G. Muhammad and M. Melhem, "Pathological voice detection and binary classification using mpeg-7 audio features," *Biomedical Signal Processing and Control*, vol. 11, pp. 1–9, 2014.

[15] J. D. Arias-Londono, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, and G. Castellanos-Domínguez, "Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients," *IEEE Transactions on biomedical engineering*, vol. 58, no. 2, pp. 370–379, 2010.

[16] A. Al-Nasheri, G. Muhammad, M. Alsulaiman, Z. Ali, T. A. Mesallam, M. Farahat, K. H. Malki, and M. A. Bencherif, "An investigation of multidimensional voice program parameters in three different databases for voice pathology detection and classification," *Journal of Voice*, vol. 31, no. 1, pp. 113–e9, 2017.

[17] R. Ranjbarzadeh, S. Dorosti, S. Jafarzadeh Ghoushchi, S. Safavi, N. Razmjooy, N. Tataei Sarshar, S. Anari, and M. Bendechache, "Nerve optic segmentation in ct images using a deep learning model and a texture descriptor," *Complex & Intelligent Systems*, vol. 8, no. 4, pp. 3543–3557, 2022.

[18] C. Yan and N. Razmjooy, "Kidney stone detection using an optimized deep believe network by fractional coronavirus herd immunity optimizer," *Biomedical Signal Processing and Control*, vol. 86, p. 104951, 2023.

[19] M. Naeem, W. K. Mashwani, M. Abiad, H. Shah, Z. Khan, and M. Aamir, "Soft computing techniques for forecasting of covid-19 in pakistan," *Alexandria Engineering Journal*, vol. 63, pp. 45–56, 2023.

[20] N. Razmjooy, V. V. Estrela, and H. J. Loschi, "Entropy-based breast cancer detection in digital mammograms using world cup optimization algorithm," in *Research Anthology on Medical Informatics in Breast and Cervical Cancer*. IGI Global, 2023, pp. 645–665.

[21] K. Shojaei and M. Abdolmaleki, "Saturated observer-based adaptive neural network leader-following control of n tractors with n-trailers with a guaranteed performance," *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 1, pp. 15–37, 2021.

[22] P. Singh, M. Sahidullah, and G. Saha, "Modulation spectral features for speech emotion recognition using deep neural networks," *Speech Communication*, vol. 146, pp. 53–69, 2023.

[23] R. Islam, M. Tarique, and E. Abdel-Raheem, "A survey on signal processing based pathological voice detection techniques," *IEEE Access*, vol. 8, pp. 66 749–66 776, 2020.

[24] H. Chen, L. Ran, X. Sun, and C. Cai, "Sw-wavenet: Learning representation from spectrogram and wavegram using wavenet for anomalous sound detection," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[25] T. Kaneko, K. Tanaka, H. Kameoka, and S. Seki, "istftnet: Fast and lightweight mel-spectrogram vocoder incorporating inverse short-time fourier transform," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 6207–6211.

[26] M. S. Khan, N. Salsabil, M. G. R. Alam, M. A. A. Dewan, and M. Z. Uddin, "Cnn-xgboost fusion-based affective state recognition using eeg spectrogram image analysis," *Scientific Reports*, vol. 12, no. 1, p. 14122, 2022.

[27] G. Aggarwal, K. Jhajharia, J. Izhar, M. Kumar, and L. Abualigah, "A machine learning approach to classify biomedical acoustic features for baby cries," *Journal of Voice*, 2023.

[28] M. Du, S. Liu, T. Wang, W. Zhang, Y. Ke, L. Chen, and D. Ming, "Depression recognition using a proposed speech chain model fusing speech production and perception features," *Journal of Affective Disorders*, vol. 323, pp. 299–308, 2023.

[29] L. Jing, M. Zhao, P. Li, and X. Xu, "A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox," *Measurement*, vol. 111, pp. 1–10, 2017.

[30] A. Abeysinghe, M. Fard, R. Jazar, F. Zambetta, and J. Davy, "Mel frequency cepstral coefficient temporal feature integration for classifying squeak and rattle noise," *The Journal of the Acoustical Society of America*, vol. 150, no. 1, pp. 193–201, 2021.

[31] S. Chachada and C.-C. J. Kuo, "Environmental sound recognition: A survey," *APSIPA Transactions on Signal and Information Processing*, vol. 3, p. e14, 2014.

[32] A. Abeysinghe, M. Fard, R. Jazar, F. Zambetta, and J. Davy, "Mel frequency cepstral coefficient temporal feature integration for classifying squeak and rattle noise," *The Journal of the Acoustical Society of America*, vol. 150, no. 1, pp. 193–201, 2021.

[33] M. S. Hossain and G. Muhammad, "Environment classification for urban big data using deep learning," *IEEE Communications Magazine*, vol. 56, no. 11, pp. 44–50, 2018.

[34] S. Tiwari, V. Sapra, and A. Jain, "Heartbeat sound classification using mel-frequency cepstral coefficients and deep convolutional neural network," in *Advances in Computational Techniques for Biomedical Image Analysis*. Elsevier, 2020, pp. 115–131.

[35] S.-H. Fang, Y. Tsao, M.-J. Hsiao, J.-Y. Chen, Y.-H. Lai, F.-C. Lin, and C.-T. Wang, "Detection of pathological voice using cepstrum vectors: A deep learning approach," *Journal of Voice*, vol. 33, no. 5, pp. 634–641, 2019.

[36] C. Vikram and K. Umarani, "Pathological voice analysis to detect neurological disorders using mfcc and svm," *Int. J. Adv. Electr. Electron. Eng*, vol. 2, no. 4, pp. 87–91, 2013.

[37] A. Sasou, "Automatic identification of pathological voice quality based on the grbas categorization," in *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 2017, pp. 1243–1247.

[38] S. Shinohara, Y. Omiya, M. Nakamura, N. Hagiwara, M. Higuchi, S. Mitsuyoshi, and S. Tokuno, "Multilingual evaluation of voice disability index using pitch rate," *Adv. Sci. Technol. Eng. Syst. J*, vol. 2, no. 3, pp. 765–772, 2017.

[39] M. Sarria-Paja and G. Castellanos-Domínguez, "Robust pathological voice detection based on component information from hmm," in *Advances in Nonlinear Speech Processing: 5th International Conference on Nonlinear Speech Processing, NOLISP 2011, Las Palmas de Gran Canaria, Spain, November 7-9, 2011. Proceedings 5*. Springer, 2011, pp. 254–261.

[40] M. R. Jamaludin, S. H. Salleh, T. T. Swee, K. Ahmad, A. K. Ibrahim, and K. Ismail, "An improved time domain pitch detection algorithm for pathological voice," *American Journal of Applied Sciences*, vol. 9, no. 1, p. 93, 2012.

[41] M. A. A. Albadr, S. Tiun, F. T. Al-Dhief, and M. A. Sammour, "Spoken language identification based on the enhanced self-adjusting extreme learning machine approach," *PloS one*, vol. 13, no. 4, p. e0194770, 2018.

[42] M. A. A. Albadr, S. Tiun, M. Ayob, F. T. Al-Dhief, K. Omar, and F. A. Hamzah, "Optimised genetic algorithm-extreme learning machine approach for automatic covid-19 detection," *PloS one*, vol. 15, no. 12, p. e0242899, 2020.