IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 14, 2017, accepted April 16, 2017, date of publication April 24, 2017, date of current version March 9, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2696056

Voice Pathology Detection and Classification Using Auto-Correlation and Entropy Features in Different Frequency Regions

AHMED AL-NASHERI[1], GHULAM MUHAMMAD[1], MANSOUR ALSULAIMAN[1], ZULFIQAR ALI[1], KHALID H. MALKI[2], TAMER A. MESALLAM[2], AND MOHAMED FARAHAT IBRAHIM[2]

[1]Digital Speech Processing Group, Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

[2]Research Chair of Voice, Swallowing, and Communication Disorders, ENT Department, College of Medicine, King Saud University, Riyadh 11543, Saudi Arabia

Corresponding author: Mansour Alsulaiman (msuliman@ksu.edu.sa)

This work was supported by the Deanship of Scientific Research at King Saud University through Research Group under Grant RG-1437-018.

ABSTRACT Automatic voice pathology detection and classification systems effectively contribute to the assessment of voice disorders, enabling the early detection of voice pathologies and the diagnosis of the type of pathology from which patients suffer. This paper concentrates on developing an accurate and robust feature extraction for detecting and classifying voice pathologies by investigating different frequency bands using autocorrelation and entropy. We extracted maximum peak values and their corresponding lag values from each frame of a voiced signal by using autocorrelation as features to detect and classify pathological samples. We also extracted the entropy for each frame of the voice signal after we normalized its values to be used as the features. These features were investigated in distinct frequency bands to assess the contribution of each band to the detection and classification processes. Various samples of the sustained vowel /a/ for both normal and pathological voices were extracted from three different databases in English, German, and Arabic. A support vector machine was used as a classifier. We also performed *u-tests* to investigate if there is a significant difference between the means of the normal and pathological samples. The best achieved accuracies in both detection and classification varied depending on the used band, method, and database. The most contributive bands in both detection and classification were between 1000 and 8000 Hz. The highest obtained accuracies in the case of detection were 99.69%, 92.79%, and 99.79% for Massachusetts eye and ear infirmary (MEEI), Saarbrücken voice database (SVD), and Arabic voice pathology database (AVPD), respectively. However, the highest achieved accuracies for classification were 99.54%, 99.53%, and 96.02% for MEEI, SVD, and AVPD, correspondingly, using the combined feature.

INDEX TERMS Voice pathology detection and classification, frequency investigation, Arabic voice pathology database (AVPD), Saarbrücken voice database (SVD), Massachusetts eye and ear infirmary (MEEI).

I. INTRODUCTION

Individuals are increasingly at risk of pathological voice problems. Around 25% of the world population whose professions compel them to speak excessively louder than the normal level suffer from these kinds of problems. For instance, singers, actors, lawyers, teachers, auctioneers, aerobics instructors, and manufacturing supervisors are all considered to work in professions that make heavy demands on the voice. As a consequence, working on the digital processing of speech signals has been found to provide a noninvasive analytical technique that is considered to be an effective assisting tool to medical doctors when identifying voice disorders, specifically in their early stages. Voice pathologies affect the vocal folds during the phonation process. They make vocal folds producing irregular vibrations due to the malfunctioning of different factors contributing to vocal vibrations. Vocal folds are differently affected by vocal fold pathologies resulting in variation in the vibratory cycle of vocal folds because their ability to be closed properly is decreased. Voice disorders also affect the shape of the vocal tract (supra-glottal) and produce irregularities in spectral properties [1]. It is well known that there is no intralaryngeal

(tracheobronchial tree) effect on the vocal tract during the production of a vowel if we consider that the voicing source has infinite resistance. However, an accurate detailed analysis must realize that the infralaryngeal structures do influence the vocal tract and that the articulatory configuration in the vocal tract interacts with the articulation in the vocal folds [38]. Hence, supplementary vocal tract-related information is predictable and thus enables the detection of the characteristics of the vocal folds, essentially during phonation [39]. In addition, voice disorders affect vocal fold vibration differently depending on the type of disorder and location of the disease in the vocal folds, making them produce different basic tones. Vocal fold vibration depends on several factors such as the mucus present on the vocal fold's tissue, stiffness, tension, muscles in the larynx, and the closing and opening of the folds. These factors are affected differently for various voice pathologies. Due to the position and size of the pathologies, the vocal folds are closed differently during vibration. Therefore, vibration varies from one type of pathology to another. This vibration produces glottal source excitation frequencies and affects the supra-glottal (the bottom part of the vocal tract) area, which in turn contributes to the frequency of the output voice signal.

Suffering from various voice pathologies like dysphonia has increased dramatically, with approximately 7.5 million people in the United States alone suffering from vocal problems [2]. However, in Saudi Arabia, around 15% of all patients who visit King Abdul Aziz University Hospital in Riyadh present with various voice disorders [3]. The influence of voice problems on teaching professionals is greater than on other professionals. Different studies in the United States have exposed that the spread of voice disorders during the whole life of people is 57.7% for teachers and 28.8% for non-teachers [4]. Further, in the Riyadh area of Saudi Arabia, around 33% of male and female teachers experience voice disorders at some point in their lives [5]. Various cases were checked by the Communication and Swallowing Disorders Unit at King Abdul Aziz University Hospital, and they found about 760 cases yearly in people with different professional and etiological backgrounds. The use of noninvasive methods (e.g., computer programs) to detect or classify pathological problems in speech has increased over time, and in the last decade various studies have been performed using automatic detection and classification of vocal fold disorders. However, these need to be investigated because of the lack of standard methods and equipment for voice disorders. The first and the most critical step to diagnose and correctly control a voice disorder is the detection of pathology. Objective assessment, including acoustical analysis, is independent of human intervention and can assist clinicians in making decisions. We firmly believe that clinicians have the final decision regarding the medical diagnosis, and an objective assessment can only be used as an assistive tool. In contrast, the subjective measurement of voice quality depends on human experience and can vary from one individual to another. There are different types of signal analyses that can be used to perform automatic voice pathology, such as long-term and short-term signal analyses. Whilst the parameters of long-term signal analyses can be taken from the acoustic analysis [6] of a speech signal or Big data [7], the parameters of short-term signal analyses can be computed by using linear predictive coefficients [8], wearable 2.0 [9], linear predictive cepstral coefficients [10], Mel-frequency cepstral coefficients [11], [12], and so on. Various machine-learning algorithms, such as Gaussian mixture model [13], [14], hidden Markov model [15], support vector machine (SVM) [16], artificial neural network [17], and so on, have been used to discriminate between pathological and normal samples. In reality, there are diverse numbers of long-term acoustic features such as pitch, shimmer, jitter, amplitude perturbation quotient, pitch perturbation quotient, harmonic-to-noise ratio, normalized noise energy, voice turbulence index, soft phonation index, frequency amplitude tremor, and glottal-to-noise excitation ratio that can be used to diagnose voice pathology as reported in [2]–[12] and [14]. For example, jitter and shimmer are appropriate features to extract vocal fold vibratory characteristics from normal and pathological samples. These parameters are widely used in other systems [18]. Further, shimmer, jitter, and seven other parameters were extracted as an iterative residual signal estimator in Rosa et al. [19], where jitter obtained 54.8% of accuracy for 21 pathologies. In addition, 33 various long-term acoustic parameters were extracted from the Multi-Dimensional Voice Program (MDVP) [20], wher Arjmandi et al. [21] provided the definition of each of these parameters and used only 22 of them. The selection of these 22 acoustic parameters was dependent on the extracted list of voice samples in the Massachusetts Eye and Ear Infirmary (MEEI) database. After the calculation for each sample (50 dysphonic and 50 normal) of these parameters, they were fed to six different classifiers for the sake of comparing their accuracies. Before applying the classification method, the authors used two distinctive reduction techniques. The best acquired accuracy for recognition is 94.26% in the case of using the binary classification SVM. Wang et al. [22], Mel-frequency cepstral coefficients and six acoustic parameters (jitter, shimmer, harmonic-to-noise ratio, soft phonation index, amplitude perturbation quotient, and relative average perturbation) were extracted, with the results compared with those of the neural network-based voice pathology detection system [23]. Sáenz-Lechón et al. compared their proposed parameters based on wavelet transform with some of the MDVP parameters to discriminate between pathological and normal samples [24]. To ensure the reliability of the acoustic MDVP parameters, some of them were compared with the same parameters extracted by using Praat. The results showed no significant difference between the two computer software approaches [25]. In recent years, different regression-based features (e.g., MPEG-7 audio descriptors and multidirectional approaches) have been used for determining voice pathologies, with high accuracy of detection [26], [27]. Another recent study investigated the most discriminative

frequency region for voice pathology detection [28]. Vowel formants were also proved to be efficient in voice pathology detection [40].

Autocorrelation function is considered to be one of the most common methods for extracting various characteristics from speech signals. Autocorrelation is known to be a domain that has certain good properties that can be used as features. When based on a correlation function applied to a short section of a voice signal, the method can provide substantial information that enables us to estimate the irregularity in the vocal folds. For example, these methods result in many peak values, with a periodicity the same as that of the input signal. Therefore, to examine the periodicity of the signal, it is common to examine its autocorrelation function. This indicates that the correlation function of a periodic signal is also periodic. Consequently, finding the pitch and fundamental frequency of the signal will be possible by using these methods. In much research, it is observed that a normal voice has more periodicities than a pathological one. Therefore, performing correlation functions on these types of classes will provide an excellent indication that it can be used to discriminate between normal and pathological voices. For instance, Von Leden, Moore, and Timke observed that pathological samples have a strong tendency for frequent and rapid changes in regularity [29]. In addition, Lieberman found that pathological voices tend to show unusually large cycle-to-cycle fluctuations in the fundamental period [30]. In this work, we perform autocorrelation on the signal frame by frame, and we also perform the entropy on the signal one frame at a time after we normalize each one. It is preferable to use a short segment of the voice signal instead of the whole signal, because the noise tends to be cancelled out in the autocorrelation process in this short segment [31].

As we observe, every voice disorder produces different frequencies depending on the type of voice disorder and its location on the vocal folds, as we described above. Consequently, observing the frequency bands is very important in order to assess which one contributes more to the detection and classification of voice disorders. For instance, Pouchoulin *et al.* [37] found that lower frequencies (0 3000 Hz) are more suitable for identifying dysphonic voices than higher frequencies. In addition, Fraile et al. [36] found that the power of dysphonic voice signal is significantly less stable in the frequency region between 2000 and 6400 Hz than the other frequency regions.

In this paper, we are trying to develop a less-expensive computational method dedicated to detecting and classifying voice pathology. Particularly, we focus on extracting features that have low dimension. In the proposed method of this study, the voice signal is fed to a bank of band pass filters, and the output of each filter will be divided into different numbers of overlapped frames. Autocorrelation function is applied one frame at a time in order to extract the peak and its corresponding lag, which will be finally represented as features. Moreover, entropy is applied to extract another feature to be stored as features. To detect and classify the voice pathology, the proposed method is evaluated by using three different databases that have three voice disorders in common: (i) the MEEI [32]; (ii) the Saarbrücken Voice Database (SVD) [33]; and (iii) the Arabic Voice Pathology Database (AVPD).

TABLE 1. Normal and pathological samples from the three databases.

Database	Normal	Pathological			
		Cysts	Paralysis	Polyps	Total
AVPD	169 (102,67)	25 (7,18)	56 (25, 31)	46 (26, 20)	127
MEEI	53 (19, 34)	10 (6, 4)	71 (39, 32)	20 (11, 9)	101
SVD	266 (130, 136)	6 (1, 5)	212 (73, 139)	45 (26, 19)	263

II. MATERIALS AND METHODS

A. DATA

In this study, we used three different databases (MEEI, SVD, and AVPD) and chose three types of voice pathologies that were common to all three of the databases — (1) vocal fold cysts; (2) unilateral vocal fold paralysis; and (3) vocal fold polyps. The number of samples in each database is shown in Table 1, where the numbers of male and female speakers are shown, respectively, inside parentheses. The three used databases are each described below.

B. MEEI VOICE DISORDER DATABASE

This database, developed by the MEEI Voice and Speech Lab, contains more than 1,400 voiced samples of the sustained vowel /a/ and the first part of the Rainbow Passage. It is commercialized by Kay Elemetrics [32] and was recorded in two different environments. The sampling frequency for normal samples was 50 kHz, while that of pathological samples was 25 kHz or 50 kHz. It is used in most studies of voice pathology detection and classification even though it has many disadvantages, such as the different environments and sample frequencies used to record normal and pathological voices. In this database, many tools were used to evaluate voice condition, including stroboscopy, acoustic aerodynamic measures, and a physical examination of the neck and mouth (this information is provided by Kay Elemetrics). In the CD provided by Kay Elemetrics, we filtered the filenames according to the three diseases; if there were multiple pathologies for a file, we ignored that file. For normal speakers, we chose all available 53 samples, and we selected only sustained vowel /a/ samples.

C. SVD

The Institute of Phonetics at Saarland University was responsible for recording the SVD database, which is freely downloadable [33]. This database contains sustained vowels/a/, /i/, and /u/ with different intonations (normal, low, high, low-high-low), along with a spoken sentence in German ''Guten Morgen, wie geht es Ihnen?'' which translates into English as ''Good morning, how are you?'' These attributes

make it a good database for researchers to use when conducting experiments. All recorded voices in the SVD were sampled at 50 kHz with 16-bit resolution. This database is new, and thus very few studies of voice pathology detection have used it. We downloaded the files from the website mentioned in [33] by using the criteria of the three diseases, and selected only the sustained vowel/a/ samples produced at the normal pitch.

D. AVPD

The voice and speech samples in this database were recorded in different sessions at the Communication and Swallowing Disorders Unit [3] of King Abdul Aziz University Hospital in Riyadh, Saudi Arabia. The recording process was performed by experienced phoneticians in a sound-treated room, and a standardized recording protocol was used to collect voices from the patient. The protocol of the database was designed to avoid the various shortcomings of the MEEI database [24]. The AVPD has recordings of sustained vowels as well as the speech of patients who have vocal fold pathologies, along with the same recordings of persons with normal speech. Normal and pathological vocal folds were determined after clinical assessment by using a laryngeal stroboscope. In the case of pathology, the perceptual severity of voice disorders was rated on a scale of 1–3, where 3 represents the most severe case. A severity rating was associated with each sample based on the consensus of a panel of three expert medical doctors. The recording has different types of texts: (1) three sustained vowels with onset and offset information; (2) isolated words, including Arabic digits and some other common words; and (3) continuous speech. The selected text was carefully chosen to cover all Arabic phonemes. Most of the speakers recorded three utterances of each vowel,/a/, /u/, and /i/. However, the isolated words and continuous speech were recorded only once in order to avoid overburdening the patients. The sample frequency of all collected normal and pathological samples in AVPD is 50 kHz. The recording process was performed by using the computerized speech lab program. The voice disorders recorded in this database were evaluated and validated by different specialist doctors at King Abdul Aziz University Hospital. Among the recorded samples, only the recordings of patients with vocal fold cysts, vocal fold polyps, and unilateral vocal fold paralysis were included in this study. We selected only sustained vowel /a/ samples.

III. PROPOSED METHOD

The key point of this study is to extract features that can enhance accuracy when detecting and classifying voice pathology, as well as to investigate the effect of different frequency regions (bands) on the detection and classification processes. In this study, we used two different methods to extract the features: autocorrelation and entropy. These features are used to discriminate between normal and pathological samples and to classify the latter. Figure 1 shows the block diagram of the proposed method, and reveals that the

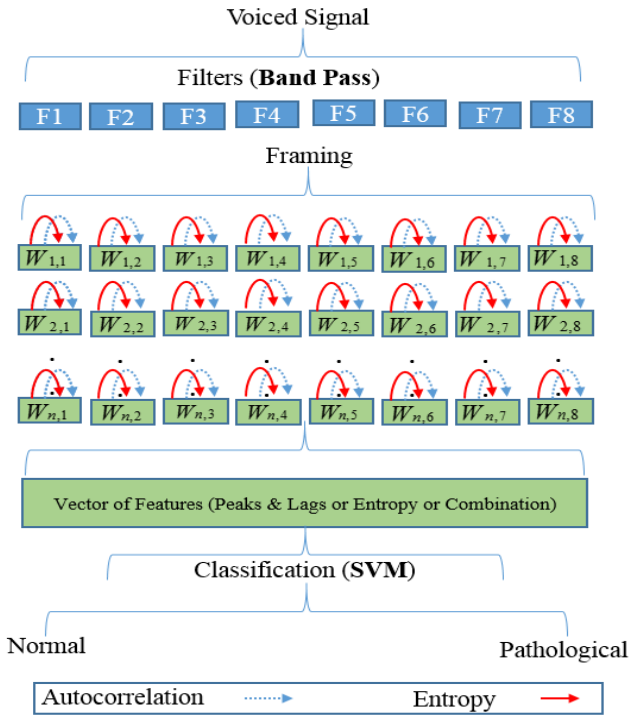

FIGURE 1. Detailed block diagram of the proposed method.

voice signal was fed to a filter bank composed of eight band pass filters. These filters represent the band pass of finite impulse response (FIR) filters, whose center frequencies are distributed on an octave scale at 31.25, 93.75, 187.5, 375, 750, 1500, 3000, and 6000 Hz. Although we also performed experiments with the Mel scale, the octave scale showed better performance in detection and classification. The main idea behind using a bank of filters was to investigate the voiced signal in different frequency regions. The output of each filter was divided into different numbers of small blocks called frames, each of which was 40 ms with an overlap of 50% (20 ms). By using the first method, we applied the autocorrelation one frame at a time in order to extract the peak and its corresponding lag for each frame in each individual filter's frame, which were finally represented as a vector of features. The autocorrelation function of a signal in a frame can be computed as follows:

$$AC(\tau) = \sum_{n=0}^{N-\tau-1} s(n)s(n+\tau) \tag{1}$$

where $0 \leq \tau \leq L-1$, AC is the autocorrelation function, s is the signal, L is the maximum lag value, N is the nu mber of samples in a frame, and τ is the lag.

In the case of using the second method to extract features by using entropy, we normalized each frame in each individual filter by subtracting the mean as computed in equation (2):

$$Normalized(e_i) = e_i - \frac{1}{N}\sum_{j=1}^{N} e_j \tag{2}$$

where N is the total number of samples in the frame, and e_i is the i^{th} element in the vector. After that, we computed the entropy for each frame by using equation (3), considering the normalized values in each frame as the probability density function for each value:

$$Entropy = -\sum_{i=0}^{N-1} P_i^* \log_2(P_i) \tag{3}$$

The values of the computed entropy for all frames of each filter represent a vector of features as in method one. We passed this vector of features to the SVM classifier to make the decision about whether the given samples were pathological or normal. Finally, we combined the extracted feature by using the two methods to represent a new set of features and passed them to the SVM classifier to detect and classify the voice pathology.

IV. EXPERIMENTAL SETUP

All samples of the sustained vowel /a/, which were taken from the three databases with the three common voice pathologies in these databases, were down sampled to 25 kHz if the sample frequency of the taken sample was more than 25 kHz, in order to make sure that all samples had the same sampling frequency. In the case of detection, we performed two different experiments on each database depending on the type of method (autocorrelation or entropy) used to extract the features. After that, we performed one extra experiment on each database by combining the extracted features (peak and lag with the entropy). Further, we performed 36 experiments for each feature on each database ($36 \times 3 \times 3 = 324$ experiments). In the case of classification, we performed six different experiments on each database depending on the type of method used to extract the features and classification type. Next, we performed three extra experiments on each database by combining the features, after which we performed 36 experiments for each feature on each database depending on the type of classification ($36 \times 3 \times 9 = 972$ experiments). Moreover, we performed six cross-database experiments on each feature (peak and lag, entropy, and the combination) depending on the classification type ($3 \times 3 \times 2 = 18$ experiments). One database was used as the trained database and the other one was used as the tested database. In addition to these experiments, we performed different experiments by using u-tests and XLSTAT software to assess the significant difference between the means of normal and pathological samples for each database separately. The output of each filter of the normal samples represents the normal class, and the output of each filter of pathological samples represents the pathological class. The null hypothesis of the u-test is ''there is no significant difference between the two classes,'' while the alternate hypothesis for this test is ''there is a significant difference between the two classes.''

V. RESULTS

The results of the performed experiments for pathology detection and classification are expressed in different terms. These terms are accuracy (ACC: the ratio between correctly detected samples and the total number of samples), sensitivity (SN: the proportion of pathological samples positively identified), specificity (SP: the proportion of normal samples negatively identified), and the area under the Receiver Operating Characteristic (ROC) curve, called the Area under Curve. These terms can be calculated by using the following distinct equations:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$SN = \frac{TP}{TP + FN} \tag{5}$$

$$SP = \frac{TN}{FP + TN} \tag{6}$$

where true negative (TN) means that the system detects a normal subject as a normal subject, true positive (TP) means that the system detects a pathological subject as a pathological subject, false negative (FN) means that the system detects a pathological subject as a normal subject, and false positive (FP) means that the system detects a normal subject as a pathological subject.

The features extracted from the three different databases need to be verified in the detection and classification processes. Hence, a number of experiments were performed to check their reliability and accuracy in both processes. To ensure accuracy, different experiments for detection and classification were performed individually for each filter (1 to 8) and their consequence combination (10 folds and 108 experiments on each database, which equal 1080 runs). From the experimental results, the acquired accuracies not only varied from one database to another, but also varied in the same database depending on the types of features (peak and lag, entropy, and their combination) being tested and the number of features used to carry out the experiments. In addition, the number of features varied from one experiment to another depending on how many filters were used when performing the experiments. For example, the number of features in each filter could be one, two, or three features depending on the type of feature, and this number increased in the case of consecutive combinations between filters (number of combined filters multiplied by one, two, or three). In the case of detection, Table 2 shows the best achieved accuracies for each database with the different types of features. As we can see from this table, the accuracies varied from one database to another for the same used feature, and these obtained accuracies also varied for the same database depending on the type of feature used to carry out the experiment. We can also notice that the combined feature had better accuracies than the other two features on the three used databases. Moreover, the performance of the used features had the best performance when using the MEEI database.

Generally, the highest achieved accuracies are 99.96%, 92.79%, and 92.79% for MEEI, SVD, and AVPD, respectively. To check the performance of each filter and combined

TABLE 2. Best detection accuracies in the three different databases using various features.

Databases	Features	SN %	SP %	ACC %
MEEI	Peak and Lag	99.58	99.73	99.67
	Entropy	99.38	99.67	99.56
	Combination	99.96	99.96	99.96
SVD	Peak and Lag	88.69	88.71	88.70
	Entropy	73.90	89.72	82.01
	Combination	91.22	94.27	92.79
AVPD	Peak and Lag	90.32	93.05	91.69
	Entropy	59.72	89.60	77.87
	Combination	91.22	94.27	92.79

filter on detection and classification, various experiments were performed, but we report only the best individual filter and best combination between consecutive filters that have the highest results. Table 3 shows the performance of the best one filter, the best combined two filters, and so on, until the best eight combined filters for each feature. As we can see from Table 3, the obtained accuracy increased in the case of the combined filters, and these accuracies differed from one database to another with the same type of feature. It is clear that the best accuracies are achieved in the case of the combination between the peak/lag features and entropy feature for the three databases. A summary of pathology detection results using the three features in differing amounts of filters from the three databases is given in Appendix A.

The highest acquired accuracies in the case of using the peak and lag feature are 99.67%, 88.70%, and 91.69% for MEEI, SVD, and AVPD, respectively, while the highest acquired accuracies in the case of using entropy are 99.56%, 82.01%, and 77.87% for MEEI, SVD, and AVPD, respectively. However, the best acquired accuracies in the case of using a combination of the two features are 99.96%, 92.79%, and 99.53% for MEEI, SVD, and AVPD, respectively. In the case of classification, we performed many different experiments depending on the used database and classification type of each feature, as shown in Table 4. The highest accuracies are achieved in the case of using the combined feature, where the classification type is "*Cyst vs (Polyp-Paralysis)*" in each database. The obtained accuracies in this case are 99.54%, 99.53%, and 96.02% for MEEI, SVD, and AVPD, respectively. In the case of using the peak and lag feature, the highest achieved accuracies are 98.72%, 99.02%, and 93.60% for MEEI, SVD, and AVPD, respectively. In addition, in the case of using the entropy feature, the highest obtained accuracies are 98.92%, 98.78%, and 85.72% for MEEI, SVD, and AVPD, respectively. As we can see in Table 4, the classification accuracies were reduced in case the classification type is *"Paralysis vs (Cyst-Polyp)"* due to the variation of this type of voice pathology, which may be unilateral or bilateral paralysis. For instance, the recorded samples in AVPD for the paralysis pathology are bilateral paralysis, whereas we are not sure about the recorded paralysis samples in MEEI and SVD (i.e., whether they are unilateral or bilateral paralysis).

Moreover, to assess the contribution of each filter and the combined filters in the classification, we performed various experiments on the three databases with three classification types: Cyst vs (Polyp-Paralysis), Paralysis vs (Cyst-Polyp), and Polyp vs (Cyst-Paralysis). We reported only the highest acquired accuracies in the individual filter and consecutive combined filter with the three features, where C = combined feature, P = peak and lag feature, and E = entropy feature (Table 5). As we can see from Table 5, in all cases the best classification accuracies are achieved using the combined feature in all databases and for all classification types. We can infer from the obtained accuracies mentioned in Table 5 that the contribution of the individual filter is less compared with the combined filters in all cases. It is clear that the combined feature has more contributions in the classification of pathological samples than the other two features. In the case of using the MEEI database, the highest obtained accuracies are 99.54%, 98.72%, and 98.92% for combined, peak and lag, and entropy, respectively, while the highest acquired accuracies in the case of using SVD are 99.53%, 98.99%, and 98.78%, respectively. In the case of using AVPD, the highest achieved accuracies are 96.02%, 93.60%, and 85.72%, respectively. As we can see from Table 5, the best accuracies for the three databases for the three features were achieved in the case of classification type "Cyst vs (Polyp-Paralysis)."

A summary of the pathology classification results using the three features in different numbers of filters on the three databases is given in Appendix B.

In addition, cross-database experiments were performed in the voice pathology samples taken from the three databases.

These experiments were performed depending on the extracted features from each database with different types of classifications. For example, Table 6 shows the results of the cross-databases in the case of using the combined feature. As we can see from this table, the highest accuracies are achieved in the case of using SVD as the trained database and the other two databases as testing. The highest acquired accuracy is 98.38% when we used SVD as the training set and the MEEI database and AVPD as the testing set. Table 7 shows the result of the cross-databases in the case of using the peak and lag feature. The highest obtained accuracy is again 98.38% in the case of using SVD as the training test and the other two databases as the testing set.

Table 8 shows the results of the cross-databases in the case of using the entropy feature. The highest archived accuracy is 98.38% in the case of SVD as the training set and the other databases as the testing set.

The reason behind using the cross-databases was to make sure that the extracted features yielded the same detection ability and to avoid unfairly over-fitting as a result of error estimation.

TABLE 3. Best detection performance for different filter numbers from the three used databases with three different features.

Features	Combined Number of Filters	MEEI			SVD			AVPD		
		SN	SP	ACC	SN	SP	ACC	SN	SP	ACC
Peak and Lag	1	96.38	99.10	98.06	65.43	67.24	66.33	70.34	93.26	81.81
	2	98.61	99.49	99.16	73.23	77.80	75.51	77.68	91.97	84.82
	3	99.15	99.66	99.47	80.50	84.74	82.62	85.89	92.64	89.26
	4	99.25	99.56	99.44	81.63	87.46	84.55	88.81	93.28	91.04
	5	99.38	99.66	99.55	85.18	88.38	86.78	86.67	92.78	89.73
	6	99.35	99.70	99.57	84.20	88.61	86.41	88.50	92.98	90.74
	7	99.40	99.65	99.55	85.82	88.67	87.24	90.32	93.05	91.69
	8	99.58	99.73	99.67	88.69	88.71	88.70	87.51	93.09	90.29
Entropy	1	85.76	94.50	91.13	42.58	77.97	60.72	16.90	95.89	64.89
	2	92.75	96.62	95.12	57.31	75.25	66.51	38.98	90.92	70.53
	3	96.02	98.60	97.61	57.92	84.27	71.43	45.57	90.15	72.67
	4	98.31	99.35	98.95	69.38	88.47	79.17	47.29	90.24	73.38
	5	98.34	99.75	99.21	72.07	89.30	80.90	53.57	90.13	75.78
	6	97.87	99.48	98.87	73.90	89.72	82.01	51.44	90.50	75.16
	7	97.87	99.48	98.86	72.89	89.61	81.46	53.07	90.28	75.68
	8	99.38	99.67	99.56	70.04	88.86	79.69	59.72	89.60	77.87
Combination	1	95.09	99.49	97.80	68.55	80.84	74.86	79.15	92.40	85.93
	2	99.77	99.92	99.86	85.17	89.79	87.55	91.98	95.91	94.00
	3	99.85	99.86	99.86	89.89	93.35	91.66	97.41	97.90	97.66
	4	99.92	99.95	99.94	90.15	93.36	91.80	98.86	98.84	98.85
	5	99.92	99.96	99.94	91.22	94.27	92.79	99.29	99.06	99.17
	6	99.96	99.96	99.96	91.35	93.96	92.69	99.58	99.31	99.44
	7	99.94	99.90	99.91	91.10	94.32	92.75	99.58	99.37	99.47
	8	99.77	99.86	99.82	89.71	93.80	91.80	99.76	99.32	99.53

Finally, we performed more additional experiments by using *u-tests* to assess the ability of the extracted features in determining normal and pathological samples for each individual database.

Table 9 shows the p-values of each feature between normal and pathological samples for each database.

Figure 2 shows the ROC curve of the extracted feature from the MEEI database. It demonstrates that the best performance is obtained in the case of the combined feature, and also shows that the performance of peak and lag is better than that of entropy.

Figure 3 also shows the ROC curve of the extracted features from AVPD with the highest obtained performance for the combined feature.

Figure 4 shows that the ROC curve of the three features extracted from SVD has the best performance for the combined feature.

From the three ROC curves mentioned above, it is clear that the extracted features from each database serve as a discriminant between normal and pathological samples. The 95% confidence interval is [0.9449 0.9870], and the one-tailed p-value is zero (<0.05), describing the significance of the data in the two classes.

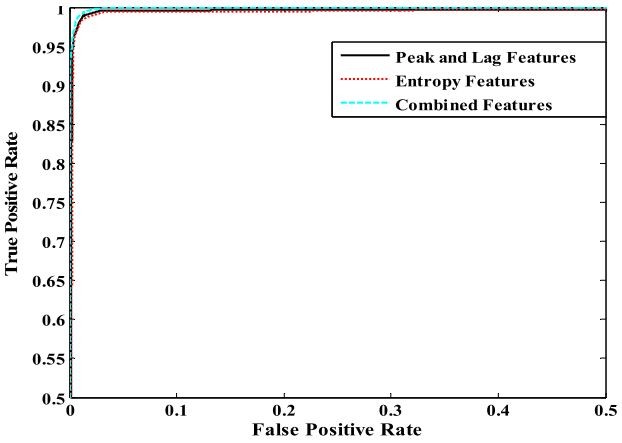

FIGURE 2. ROC curve for MEEI by using the three features.

VI. DISCUSSION

In this study, two different methods were used to extract the features, as previously discussed. We investigated

TABLE 4. Best accuracies for classification for the three databases with the three features for a different number of filters.

Classification Type	Number of Filters	MEEI Accuracy %			SVD Accuracy %			AVPD Accuracy %		
		C	P	E	C	P	E	C	P	E
Cyst vs (Paralysis-Polyp)	1	94.89	92.13	90.08	98.57	98.38	98.38	88.68	84.49	81.69
	2	98.51	96.54	92.26	99.01	98.44	98.42	93.85	88.89	81.99
	3	99.13	98.70	96.29	99.48	99.02	98.46	96.02	91.98	83.45
	4	99.34	98.72	97.60	99.43	98.99	98.78	95.78	92.67	85.57
	5	99.44	98.34	97.14	99.41	98.73	98.39	95.92	92.98	85.21
	6	99.54	98.49	97.25	99.37	98.48	98.49	95.84	93.60	85.03
	7	99.32	98.39	98.92	99.53	98.98	98.38	95.59	92.34	85.72
	8	99.54	97.74	97.91	99.42	98.38	98.53	93.69	91.15	85.47
Paralysis vs (Polyp-Cyst)	1	90.27	85.45	72.05	85.20	82.08	82.34	80.36	74.93	59.24
	2	97.18	94.62	78.39	90.19	84.64	82.40	89.04	82.68	64.87
	3	97.72	96.63	89.28	94.71	87.09	82.94	91.32	87.23	72.08
	4	97.78	96.34	96.31	94.80	90.54	83.28	90.95	87.94	73.59
	5	97.06	95.94	94.49	94.68	91.12	85.10	90.73	88.23	75.57
	6	98.07	95.47	94.60	94.83	90.42	83.46	91.34	88.76	75.27
	7	97.83	95.40	95.38	94.34	89.97	83.72	90.87	88.51	77.68
	8	97.00	94.18	94.70	94.70	90.32	84.15	91.37	88.32	76.37
Polyp vs (Cyst-Paralysis)	1	93.98	90.08	80.54	87.44	83.31	83.95	81.69	76.27	65.64
	2	97.72	94.80	82.84	91.00	86.33	84.06	89.45	86.27	69.30
	3	98.20	96.23	92.71	93.34	89.66	84.50	91.76	89.73	76.73
	4	98.18	96.54	96.87	95.44	90.15	84.95	92.65	90.34	77.33
	5	98.45	96.73	97.47	95.13	90.98	85.48	92.60	90.63	78.13
	6	98.36	97.10	95.73	94.59	89.97	85.55	92.09	91.41	79.74
	7	98.07	96.73	95.98	92.67	90.12	86.61	90.57	90.35	77.75
	8	97.78	95.34	96.23	94.68	90.96	85.13	91.45	88.21	78.63

the extracted features and their combination for different frequency regions for voice pathology detection and classification. From the experimental results mentioned above, we found a variation in the obtained accuracies in the same database. This variation relates to the different types of used features and different bands of each filter. In the case of the used features, the reason behind these variations is that each feature has different values than the other features. In the case of the different filters, the variation in the accuracies in the same database also differed from one filter to another, as the frequency bands of each filter were different, which indicates that every frequency band has a different contribution to the detection and classification of pathologies. Figure 5 reflects this variation in the accuracies in the case of the extracted features from the MEEI database. As it is seen from this figure, the greatest contribution for detection and classification is achieved in the case of filters 4, 5, 6, and 7. This also confirms the findings of Fraile et al. [36], who state that "the power of dysphonic voices' signal is significantly less stable in the frequency region between 2000 and 6400 Hz than the other frequency regions." Further, as we notice from the experimental results, the highest achieved accuracies occurred in the combined filters in all experiments. This is because each filter has the ability to detect some components in the specified range of frequencies than another filter. Moreover, when we combine more than one filter, their frequency range is expanded, which leads to higher accuracy detection and classification.

The variation of the obtained accuracy from one database to another may be caused by different reasons: (1) the severity

TABLE 5. Best classification accuracies (%) for the three used databases with three different types of classifications for the three features.

Databases	Classification Type	Features		
		Combination	Peak and Lag	Entropy
MEEI	Cyst vs (Polyp-Paralysis)	99.54	98.72	98.92
	Paralysis vs (Cyst-Polyp)	98.07	96.63	96.31
	Polyp vs (Cyst-Paralysis)	98.45	97.10	97.47
SVD	Cyst vs (Polyp-Paralysis)	99.53	99.02	98.78
	Paralysis vs (Cyst-Polyp)	94.83	91.12	85.10
	Polyp vs (Cyst-Paralysis)	95.44	90.98	86.61
AVPD	Cyst vs (Polyp-Paralysis)	96.02	93.60	85.72
	Paralysis vs (Cyst-Polyp)	91.37	88.76	77.68
	Polyp vs (Cyst-Paralysis)	92.65	91.41	79.74

TABLE 6. Cross-database experimental results in the case of using the combined feature.

Cross-Database Accuracies				
Training	Classification Type	Test		
		MEEI	SVD	AVPD
MEEI	Cyst vs (Paralysis & Polyp)		90.07	90.07
	Paralysis vs (Cyst & Polyp)		70.24	52.68
	Polyp vs (Cyst & Paralysis)		80.17	80.17
SVD	Cyst vs (Paralysis & Polyp)	98.38		98.38
	Paralysis vs (Cyst & Polyp)	81.37		66.41
	Polyp vs (Cyst & Paralysis)	82.99		82.99
AVPD	Cyst vs (Paralysis & Polyp)	81.20	81.20	
	Paralysis vs (Cyst & Polyp)	46.59	47.48	
	Polyp vs (Cyst & Paralysis)	65.39	64.94	

of voice pathologies, which are not the same for the three databases, as shown in Table 2 for instance, where sensitivity (to pathological samples) varies from one database to another; (2) the recording environment and regulation of the recording, which are not the same for the three databases; (3) in the case of the MEEI database, the recording environments for pathological and normal samples are not the same; and (4) the number of samples taken from each database in this study are not the same. The variation of the accuracies in the three different databases is illustrated in Figure 6. It is obvious from this figure that the highest accuracies for the eight different bands are acquired in the case of using the feature that was extracted from the MEEI database.

In addition, the obtained results from the cross-database shown in Tables 6, 7, and 8 indicate that the best accuracies

TABLE 7. Cross-database experimental results in the case of using the peak and lag feature.

Cross-Database Accuracies				
Training	Classification Type	Test		
		MEEI	SVD	AVPD
MEEI	Cyst vs (Paralysis & Polyp)		90.07	90.07
	Paralysis vs (Cyst & Polyp)		70.24	46.20
	Polyp vs (Cyst & Paralysis)		80.17	80.17
SVD	Cyst vs (Paralysis & Polyp)	98.38		98.38
	Paralysis vs (Cyst & Polyp)	81.37		59.84
	Polyp vs (Cyst & Paralysis)	82.99		82.99
AVPD	Cyst vs (Paralysis & Polyp)	81.20	81.20	
	Paralysis vs (Cyst & Polyp)	46.59	46.59	
	Polyp vs (Cyst & Paralysis)	65.39	65.39	

TABLE 8. Cross-database experimental results in the case of using the entropy feature.

Cross-Database Accuracies				
Training	Classification Type	Test		
		MEEI	SVD	AVPD
MEEI	Cyst vs (Paralysis & Polyp)		90.07	90.07
	Paralysis vs (Cyst & Polyp)		70.24	31.69
	Polyp vs (Cyst & Paralysis)		80.17	80.17
SVD	Cyst vs (Paralysis & Polyp)	98.38		98.38
	Paralysis vs (Cyst & Polyp)	81.37		22.05
	Polyp vs (Cyst & Paralysis)	82.99		82.99
AVPD	Cyst vs (Paralysis & Polyp)	81.20	81.20	
	Paralysis vs (Cyst & Polyp)	46.59	47.65	
	Polyp vs (Cyst & Paralysis)	65.39	64.92	

were obtained in the case of using SVD as the training set and the other databases as the testing set. It is also seen from these tables that the obtained accuracies varied from one classification type to another. The best acquired accuracy was in the case of the "Cyst vs (Paralysis & Polyp)" classification type, while the worst obtained accuracy was in the case of "Paralysis vs (Cyst & Polyp)" for all the different features. The reason behind the decrease in the achieved accuracies in the case of "Paralysis vs (Cyst & Polyp)" is that this type of voice pathology comes with two different types: unilateral or bilateral paralysis. In the case of using AVPD, the types of recorded voices of paralysis were bilateral paralysis, whereas in the case of using MEEI and SVD, we are not sure whether the type of paralysis is unilateral or bilateral. From Tables 6, 7, and 8, we can notice that the obtained accuracies did not change much in all databases, which indicates that the features are independent of the used databases.

TABLE 9. U-test for the three extracted features for the three used databases.

Filter No.	P-Values of MEEI Features			P-Values of SVD Features			P-Values of AVPD Features		
	Entropy	Peak	Lag	Entropy	Peak	Lag	Entropy	Peak	Lag
1	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.8015	< 0.0001	< 0.0001	< 0.0001
2	< 0.0001	< 0.0001	< 0.0001	0.0391	< 0.0001	0.7777	< 0.0001	< 0.0001	< 0.0001
3	< 0.0001	< 0.0001	< 0.0001	0.0998	< 0.0001	0.8107	< 0.0001	< 0.0001	< 0.0001
4	< 0.0001	< 0.0001	< 0.0001	0.2135	< 0.0001	0.8895	< 0.0001	< 0.0001	< 0.0001
5	< 0.0001	< 0.0001	< 0.0001	0.4876	< 0.0001	0.9367	< 0.0001	< 0.0001	< 0.0001
6	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.9442	< 0.0001	< 0.0001	< 0.0001
7	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
8	< 0.0001	< 0.0001	0.3255	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.4628

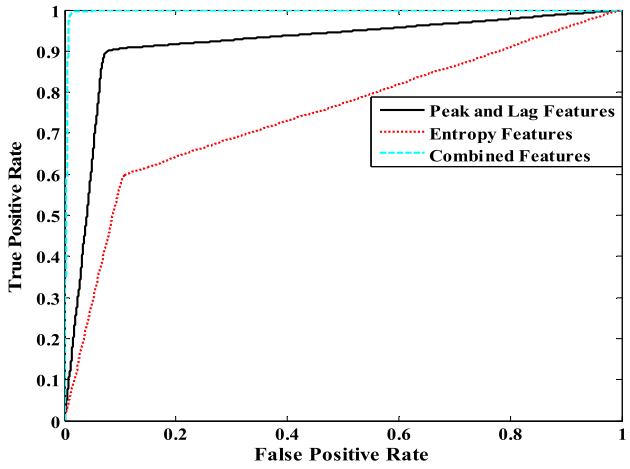

FIGURE 3. ROC curve for AVPD by using the three features.

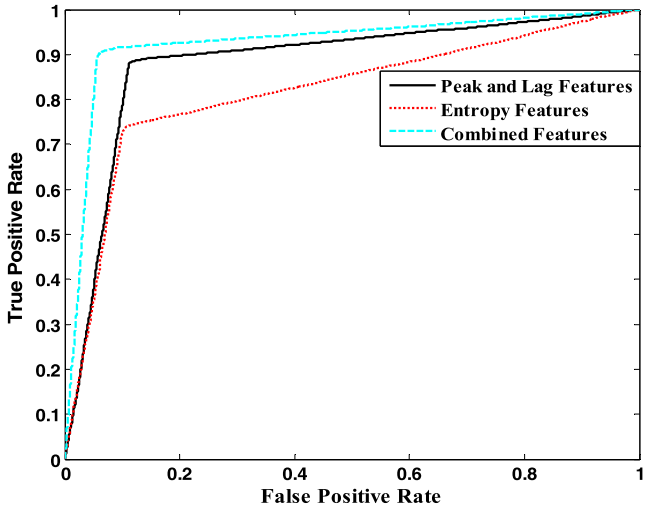

FIGURE 4. ROC curve for SVD by using the three features.

Moreover, the calculated p-values for the entropy, peak, and lag features shown in Table 9 indicate that each feature helped discriminate between normal and pathological

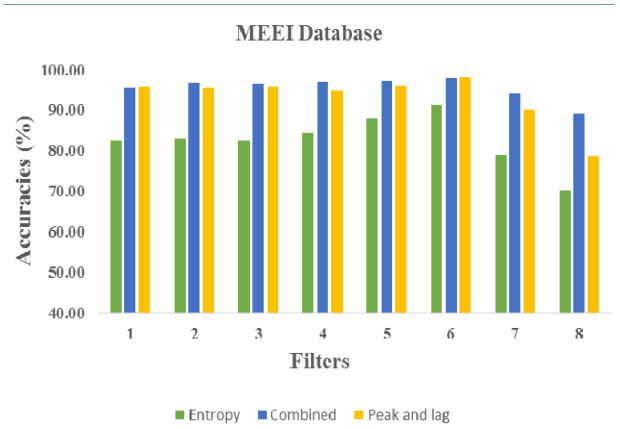

FIGURE 5. Performance of the three extracted features on each filter.

TABLE 10. Comparison of accuracies between methods (pathology detection).

Methods	MEEI	SVD	AVPD
Our Method	99.96%	92.79%	99.79%
Method [14]	94.07%	-	-
Method [21]	94.24%		-
Method [26]	99.41%	-	-
Method [34]	94.80%	81.00%	-
Method [35]	94.10%		
Method [41]	88.21%	99.68%	72.53%

samples. We can infer from this table that the peak value has the best contribution than the other two features for the three databases. On the other hand, the entropy and lag features have the best contribution in MEEI and AVPD, but their contributions were reduced in the case of SVD. In general, we can conclude from this table that the best contribution for

TABLE 11. Pathology detection results using the three features in different number of filters on the three databases.

Detection Summary																										
Serial	Filters	Entropy									Combined (Entropy + (Peak and Lag))									Peak and Lag								
		MEEI			SVD			AVPD			MEEI			SVD			AVPD			MEEI			SVD			AVPD		
		SN	SP	AC	SN	SP	AC	SN	SP	AC	SN	SP	AC	SN	SP	AC	SN	SP	AC	SN	SP	AC	SN	SP	AC	SN	SP	AC
1	1	61.52	95.66	82.51	28.49	81.16	55.50	16.90	95.89	64.89	91.85	97.80	95.52	68.55	80.84	**74.86**	76.19	92.37	84.48	90.36	98.96	95.65	61.85	65.61	63.73	70.35	93.21	81.78
2	2	65.59	93.90	83.00	31.11	78.08	55.19	15.25	96.50	64.62	93.31	98.91	96.75	66.61	79.28	73.11	79.15	92.40	85.93	90.52	98.82	95.63	57.03	67.98	62.50	70.34	93.26	81.81
3	3	62.11	95.36	82.56	27.61	81.85	55.41	16.98	95.81	64.88	92.65	98.67	96.33	67.97	79.56	73.92	77.48	92.09	84.96	90.63	98.90	95.72	61.29	65.71	63.49	70.09	93.42	81.76
4	4	70.22	93.24	84.39	30.65	78.84	55.36	16.11	96.20	64.78	93.98	98.82	96.96	65.10	78.94	72.20	75.48	92.44	84.17	89.36	98.30	94.86	60.82	66.22	63.51	72.27	91.28	81.78
5	5	78.43	94.08	88.07	40.03	70.10	55.45	17.43	95.17	64.66	94.55	98.77	97.15	63.31	78.09	70.89	78.18	93.23	85.88	91.33	98.85	95.96	59.62	61.14	60.37	70.31	93.10	81.70
6	6	85.76	94.50	91.13	58.17	57.16	57.64	16.57	94.96	64.20	95.09	99.49	97.80	63.20	79.39	71.51	75.68	93.38	84.74	96.38	99.10	98.06	65.43	67.24	66.33	71.27	91.56	81.42
7	7	70.79	83.86	78.84	24.90	80.25	53.25	0.00	100.00	60.76	89.77	96.85	94.12	61.13	86.80	74.30	73.91	86.18	80.19	84.93	93.38	90.13	57.49	71.81	64.64	73.72	82.04	77.89
8	8	29.07	96.07	70.25	42.58	77.97	60.72	1.66	99.68	61.21	83.42	92.65	89.09	59.08	78.23	68.90	70.73	77.84	74.36	54.10	94.08	78.69	56.13	69.21	62.67	72.61	68.50	70.55
9	12	92.21	96.56	94.89	68.84	60.77	64.71	17.91	95.54	65.07	94.63	98.91	97.27	76.59	83.96	80.37	87.81	95.11	91.55	97.38	99.39	98.62	71.94	67.14	69.54	72.71	92.58	82.63
10	23	92.45	96.40	94.88	68.89	60.71	64.70	23.93	94.00	66.51	98.98	99.65	99.39	75.96	84.88	80.54	86.19	94.80	90.60	98.35	99.55	99.08	69.98	77.00	73.48	72.70	91.87	82.28
11	34	92.75	96.62	95.12	58.32	72.27	65.46	34.51	90.98	68.82	99.09	99.62	99.42	79.12	86.36	82.83	86.57	95.11	90.95	98.17	99.37	98.91	73.23	77.80	75.51	76.74	91.37	84.05
12	45	93.04	95.85	94.76	52.17	76.06	64.42	38.73	90.33	70.08	99.15	99.66	99.47	80.88	87.43	84.24	92.28	94.90	93.63	98.61	99.49	99.16	68.40	75.19	71.79	73.32	92.48	82.90
13	56	87.80	97.07	93.48	52.78	68.37	60.76	38.98	90.92	70.53	99.77	99.92	99.86	85.17	89.79	**87.55**	91.98	95.91	94.00	98.39	99.55	99.10	71.67	72.48	72.08	77.68	91.97	84.82
14	67	88.78	95.36	92.82	44.00	79.80	62.36	22.33	93.39	65.51	99.46	99.66	99.59	84.32	89.65	87.05	87.51	94.63	91.15	96.72	99.43	98.38	67.38	79.09	73.23	78.00	91.55	84.77
15	78	82.55	90.35	87.31	57.31	75.25	66.51	10.47	96.58	62.79	95.90	98.15	97.28	76.05	89.69	83.05	76.40	89.03	82.87	93.12	96.44	95.16	68.44	82.07	75.26	76.32	89.64	82.98
16	123	92.52	96.52	94.98	69.03	60.62	64.72	28.99	91.98	67.25	99.30	99.69	99.54	77.97	85.23	81.69	96.40	97.64	97.04	98.21	99.48	99.00	72.94	76.84	74.89	73.18	92.68	82.92
17	234	92.76	96.51	95.06	61.96	68.18	65.16	35.79	90.43	68.99	99.19	99.69	99.50	78.26	85.94	82.20	93.63	96.19	94.94	98.34	99.45	99.03	72.16	76.63	74.41	77.47	91.44	84.46
18	345	92.62	96.27	94.86	58.25	79.68	69.24	41.05	89.93	70.74	99.17	99.69	99.49	81.79	88.18	85.07	92.18	96.27	94.28	98.45	99.47	99.08	75.88	78.57	77.22	77.67	91.05	84.35
19	456	93.62	98.85	96.84	57.92	84.27	71.43	45.57	90.15	72.67	99.34	99.90	99.68	88.38	92.15	90.32	96.92	98.01	97.48	99.15	99.66	99.47	79.17	81.25	80.21	78.51	92.25	85.38
20	567	93.25	97.62	95.93	53.17	83.68	68.81	43.11	90.79	72.07	99.85	99.86	99.86	89.89	93.35	**91.66**	97.41	97.90	97.66	98.55	99.53	99.16	80.50	84.74	82.62	85.89	92.64	89.26
21	678	96.02	98.60	97.61	59.58	81.25	70.70	36.68	90.72	69.51	99.29	99.75	99.58	87.07	90.93	89.05	89.26	94.88	92.13	97.60	99.36	98.69	72.93	85.23	79.08	82.02	91.38	86.70
22	1234	92.63	96.34	94.91	58.94	72.30	65.79	36.26	90.41	69.16	98.46	99.60	99.16	80.22	86.11	83.24	98.86	98.84	98.85	98.12	99.36	98.88	69.61	75.13	72.37	78.56	92.01	85.29
23	2345	92.48	96.25	94.80	57.83	79.79	69.08	41.37	89.69	70.73	99.00	99.61	99.38	80.15	86.99	83.66	98.16	98.67	98.42	98.38	99.45	99.04	75.41	77.64	76.53	79.90	91.48	85.68
24	3456	94.99	99.03	97.47	57.96	85.12	71.89	45.59	90.23	72.72	99.28	99.87	99.64	87.16	91.67	89.47	98.32	98.76	98.55	98.67	99.61	99.25	81.73	83.97	82.85	81.58	93.00	87.29
25	4567	96.52	99.23	98.19	63.21	85.64	74.71	47.29	90.24	73.38	99.92	99.95	99.94	90.15	93.36	**91.80**	97.93	98.36	98.15	98.90	99.61	99.34	81.63	87.46	84.55	88.81	93.28	91.04
26	5678	98.31	99.35	98.95	69.38	88.47	79.17	46.90	90.22	73.22	99.94	99.90	99.91	88.35	93.08	90.77	97.31	98.33	97.83	99.25	99.56	99.44	79.51	85.01	82.27	81.11	90.97	86.05
27	12345	92.54	97.73	95.73	58.32	80.66	69.78	41.05	89.71	70.61	98.83	99.64	99.32	83.66	88.67	86.22	98.89	99.00	98.94	98.07	99.33	98.84	72.73	76.73	74.72	77.71	91.37	84.53
28	23456	94.87	98.95	97.38	57.34	83.95	70.98	48.23	90.26	73.77	99.77	99.94	99.87	88.27	92.16	90.26	98.89	99.13	99.01	98.82	99.58	99.29	81.62	83.52	82.57	85.32	92.68	89.00
29	34567	96.68	99.19	98.22	65.55	85.90	75.98	49.23	90.29	74.17	99.92	99.97	99.95	89.82	93.49	91.71	99.29	99.06	99.17	98.50	99.55	99.14	81.02	86.58	83.79	86.67	92.78	89.73
30	45678	98.34	99.75	99.21	72.07	89.30	**80.90**	53.57	90.13	75.78	99.92	99.96	99.94	91.22	94.27	**92.79**	98.70	98.75	98.73	99.38	99.66	99.55	85.18	88.38	86.78	84.13	91.97	88.05
31	123456	94.95	98.95	97.41	57.95	85.14	71.89	46.89	90.35	73.29	99.08	99.75	99.50	88.55	92.24	90.44	99.58	99.31	99.44	98.91	99.69	99.39	82.31	84.25	83.28	82.70	93.24	87.97
32	234567	96.71	99.24	98.27	61.74	85.68	74.01	47.23	90.38	73.44	99.92	99.90	99.90	91.35	93.96	**92.69**	99.49	99.36	99.42	98.98	99.73	99.44	84.20	88.61	86.41	88.50	92.98	90.74
33	345678	97.87	99.48	98.87	73.90	89.72	**82.01**	51.44	90.50	75.16	99.96	99.96	99.96	89.74	93.28	91.56	99.44	99.10	99.27	99.35	99.70	99.57	81.86	87.97	84.92	84.99	92.24	88.62
34	1234567	96.66	99.25	98.25	62.37	85.16	74.05	50.11	90.56	74.68	99.94	99.90	99.91	90.74	93.72	92.27	99.58	99.37	99.47	98.69	99.65	99.28	85.17	89.03	87.10	90.32	93.05	91.69
35	2345678	97.87	99.48	98.86	72.89	89.61	**81.46**	53.07	90.28	75.68	99.96	99.86	99.90	91.10	94.32	**92.75**	99.61	99.23	99.41	99.40	99.65	99.55	85.82	88.67	87.24	90.00	92.67	91.33
36	12345678	99.38	99.67	99.56	70.04	88.86	79.69	59.72	89.60	77.87	99.77	99.86	99.82	89.71	93.80	**91.80**	99.76	99.32	99.53	99.58	99.73	99.67	88.69	88.71	88.70	87.51	93.09	90.29

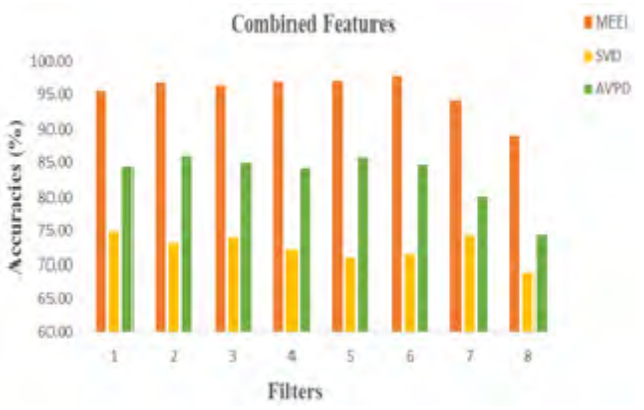

FIGURE 6. Performance of the combined features from the three used databases on each filter.

these features is peak, entropy, and lag in that order. From the experimental results, we found the following:

- Every extracted feature contributed to the classification and detection processes, but the combined feature performed better than the other features.
- The performance of the individual filter (frequency band) varied, and the best performance was in the case of using filters 4, 5, 6, and 7 (with frequency band range 1000∼8000 Hz).
- The performance of consecutive combined filters was better than that of the individual filter in the classification and detection processes.
- The used features were independent of the databases.
- The severity and variation of voice disorders were not addressed in this study and are thus left for future work.

Generally, the obtained results in this study are better than, or comparable with, other reported results using MEEI, SVD, and AVPD. For instance, the acquired accuracies were 94.26% and 91.55% for detection and classification, respectively in [21], where Arjmandi et al. used the MEEI database and SVM classifier but different pathological samples. In addition, Godino-Llorente et al. [14] used the MEEI database and reported an accuracy of 94.07%, while Martínez et al. [34] used the SVD and same classifier that we used and attained an accuracy of 81%. The authors also used the MEEI database and the acquired accuracy was 94.80%. Furthermore, Markaki et al. [35] used the MEEI database and the attained accuracy for the classification was 94.10%, while Muhammad and Melhem [26] used the MPEG feature to detect and classify the voice pathology by using

TABLE 12. Pathology classification results using the three features in different number of filters on the three databases.

Classification on the three Databases																										
No.	Filters	MEEI									SVD										AVPD					
		polyp-(cyst-paralysis)			paralysis - (polyp-cyst)			cyst-(paralysis -polyp)			polyp-(cyst-paralysis)			paralysis - (polyp-cyst)			cyst-(paralysis -polyp)			polyp-(cyst-paralysis)			paralysis - (polyp-cyst)			cyst-(paralysis -polyp)		
		CAC	PAC	EAC	CAC	PAC	EAC	CAC	PAC	EAC	CAC	PAC	EAC	CAC	PAC	EAC	CAC	PAC	EAC	CAC	PAC	EAC	CAC	PAC	EAC	CAC	PAC	EAC
1	1	92.90	83.50	80.19	83.25	81.74	70.27	93.42	90.06	90.08	85.53	83.31	83.77	84.88	81.37	82.19	98.57	98.38	98.38	81.24	75.82	65.39	79.54	74.89	58.57	87.29	83.08	81.69
2	2	93.85	90.08	80.19	90.27	85.45	70.21	93.64	92.13	90.08	86.18	82.99	82.99	83.59	82.07	81.37	98.38	98.38	98.38	81.21	76.27	65.39	78.91	74.38	58.88	88.68	82.93	81.54
3	3	92.96	89.09	80.14	86.05	82.07	70.23	94.10	91.20	90.08	87.44	83.09	82.99	85.20	81.37	81.37	98.38	98.38	98.38	81.69	76.05	65.39	80.36	74.93	59.19	87.98	83.38	81.58
4	4	93.98	80.12	80.17	85.88	83.17	70.21	94.68	90.52	90.06	87.41	82.99	82.99	84.04	81.66	81.37	98.39	98.38	98.38	81.14	76.24	65.39	80.01	74.85	58.83	88.21	84.49	81.58
5	5	92.69	84.55	80.17	88.72	79.07	70.25	94.89	90.10	90.08	86.73	83.09	83.02	83.99	82.08	81.37	98.44	98.38	98.38	80.70	75.34	65.39	79.61	74.61	59.24	87.79	82.58	81.55
6	6	91.22	84.24	80.54	87.27	79.92	70.97	94.39	90.04	90.06	85.08	82.99	83.82	83.00	81.37	82.19	98.53	98.38	98.38	80.61	75.27	65.39	78.21	73.06	58.16	86.40	82.67	81.20
7	7	86.46	84.43	80.14	81.70	80.17	72.05	92.59	90.10	90.08	85.63	82.97	83.95	84.38	81.29	82.34	98.38	98.38	98.38	74.10	73.66	65.64	66.65	66.80	53.41	81.20	81.16	81.20
8	8	82.05	81.33	80.17	71.99	71.33	70.23	90.33	90.14	90.06	83.84	82.89	83.20	81.61	81.34	81.58	98.38	98.38	98.38	69.35	69.65	65.39	64.14	64.39	53.41	81.20	81.20	81.20
9	12	93.02	91.45	80.19	91.88	88.03	70.21	94.76	93.04	90.08	87.69	83.78	83.93	84.97	82.14	82.32	98.59	98.38	98.38	83.34	77.85	65.45	80.74	76.63	60.56	88.47	83.78	81.60
10	23	95.22	91.33	80.14	92.59	91.16	70.23	95.69	90.10	90.06	86.71	84.62	83.91	85.87	83.36	82.14	98.60	98.38	98.42	82.08	81.59	65.84	81.10	77.56	60.20	88.61	85.38	81.53
11	34	95.71	94.80	80.37	89.44	90.08	70.21	95.47	94.80	90.06	87.98	83.98	83.68	87.13	83.12	82.15	98.59	98.38	98.38	84.49	81.43	66.15	84.10	79.28	60.72	90.34	85.49	81.67
12	45	97.08	91.33	80.68	94.80	93.29	72.40	94.37	94.14	89.98	90.23	85.02	83.19	87.25	81.80	81.60	98.38	98.38	98.38	85.18	85.30	68.51	85.32	78.80	64.87	92.39	85.14	81.99
13	56	97.72	91.80	82.59	97.18	94.62	75.47	97.81	96.54	90.25	91.00	86.33	83.81	90.19	84.46	82.34	98.38	98.38	98.38	89.45	86.27	69.30	89.04	82.68	63.47	93.85	88.89	81.91
14	67	96.54	90.66	82.84	95.59	90.95	78.39	98.51	95.13	92.19	88.16	85.63	83.96	89.81	84.64	82.34	99.01	98.44	98.38	86.09	84.02	65.57	81.45	81.03	59.23	93.02	83.78	81.20
15	78	92.07	81.86	80.19	90.77	78.28	77.41	95.24	94.37	92.26	83.95	85.41	84.06	82.35	81.37	82.40	98.38	98.33	98.38	76.83	76.89	65.64	76.60	74.46	53.41	83.25	81.20	81.20
16	123	94.89	91.12	80.19	91.59	90.99	70.25	95.53	92.92	90.06	88.04	83.87	83.92	87.08	81.61	82.35	98.38	98.38	98.38	82.72	81.35	66.29	81.11	80.98	60.32	89.65	83.10	81.65
17	234	96.11	94.31	80.21	94.51	91.68	70.21	95.55	94.51	90.08	88.33	84.25	83.91	86.16	82.38	82.38	98.38	98.38	98.38	84.04	85.74	65.83	84.14	80.94	59.92	91.04	84.62	81.50
18	345	97.23	93.83	80.72	95.59	90.81	72.34	97.16	96.50	90.06	90.65	83.70	84.18	88.75	84.78	82.49	98.38	98.38	98.38	85.48	88.58	74.74	85.63	83.99	69.35	93.71	85.80	81.58
19	456	98.20	94.47	86.15	97.72	94.84	83.44	98.96	96.56	91.68	91.67	85.92	84.31	93.22	87.09	82.39	99.11	98.38	98.42	91.76	88.21	76.73	89.78	84.42	72.08	95.51	88.99	83.45
20	567	97.10	96.23	92.71	97.49	96.63	89.28	99.13	98.70	96.29	93.34	89.66	83.82	94.71	86.94	82.34	99.48	99.02	98.46	91.33	89.73	71.58	91.32	87.23	65.90	96.02	91.98	81.86
21	678	96.63	93.23	91.47	94.27	88.88	87.41	98.14	95.59	94.24	88.91	87.71	84.50	91.38	86.00	82.94	99.37	98.47	98.38	84.85	80.60	65.60	86.03	82.87	61.84	92.93	88.77	81.29
22	1234	96.42	93.87	80.19	92.01	90.87	70.23	96.09	93.60	90.06	86.91	83.32	84.09	88.32	81.64	82.46	98.58	98.38	98.43	83.82	84.80	67.04	83.02	81.55	65.12	90.35	86.86	81.61
23	2345	97.06	94.14	80.19	95.51	90.93	75.49	97.72	94.76	90.06	88.92	86.25	84.20	88.89	85.28	82.53	98.78	98.38	98.38	87.52	87.79	76.14	86.01	83.17	69.84	90.95	88.20	81.90
24	3456	98.18	95.32	87.62	97.64	94.87	77.91	99.19	97.76	90.06	93.74	87.37	84.20	92.85	88.70	82.55	99.35	98.38	98.38	90.85	89.80	77.33	89.92	84.77	72.12	95.78	89.15	83.44
25	4567	98.16	96.54	93.50	97.78	96.34	90.14	99.34	98.72	97.60	95.44	90.15	84.51	94.48	90.36	82.47	98.93	98.48	98.43	92.65	90.34	77.11	90.95	87.94	73.59	95.44	92.67	85.57
26	5678	97.60	92.46	96.87	97.10	94.20	96.31	98.39	98.16	96.56	94.74	89.97	84.95	94.80	90.54	83.28	99.43	98.99	98.78	90.94	89.36	71.40	90.88	87.86	71.67	95.52	91.18	83.76
27	12345	97.16	93.83	80.87	94.72	92.67	73.23	96.34	96.98	90.12	89.35	85.03	84.21	90.16	84.04	82.57	98.38	98.38	98.39	87.41	88.31	75.69	85.20	83.85	70.82	93.17	88.63	82.06
28	23456	98.45	95.13	81.45	96.98	94.37	81.59	98.96	97.12	90.46	93.66	88.93	84.44	92.66	86.12	82.50	99.29	98.38	98.38	89.54	86.68	77.26	89.56	86.00	72.57	95.33	91.79	83.73
29	34567	98.14	96.73	94.80	96.73	95.94	91.14	99.44	98.34	95.78	95.13	88.89	84.74	93.11	90.12	83.40	99.33	98.54	98.38	92.60	90.63	78.13	90.73	88.23	75.57	95.92	92.90	85.21
30	45678	98.01	94.62	97.47	97.06	95.22	94.49	99.34	97.74	97.14	94.53	90.98	85.48	94.68	91.12	85.10	99.41	98.73	98.38	89.02	89.49	77.41	89.91	86.16	75.40	95.60	92.98	84.59
31	123456	98.36	96.42	88.01	97.64	93.31	81.12	99.11	97.81	92.63	92.40	86.63	84.08	93.34	86.94	82.85	99.22	98.47	98.35	91.25	88.99	77.43	89.05	85.51	74.09	95.75	89.98	81.89
32	234567	97.99	97.10	94.74	98.07	95.34	89.19	99.54	98.49	96.92	94.01	89.97	84.80	93.67	90.26	82.62	98.71	98.48	98.38	92.09	91.41	78.25	91.34	88.76	73.82	95.84	93.60	83.34
33	345678	98.36	94.47	95.73	97.25	95.47	94.60	99.32	98.43	97.25	94.59	88.51	85.55	94.83	90.42	83.46	99.37	98.48	98.49	90.29	88.42	79.74	89.77	88.72	75.27	93.54	92.25	85.03
34	1234567	97.99	96.73	90.99	96.85	95.40	89.50	99.32	98.24	96.40	92.67	88.88	84.79	93.97	88.42	83.02	99.34	98.46	98.38	90.57	90.35	77.75	90.87	86.92	74.99	95.59	92.34	85.55
35	2345678	98.07	96.44	95.98	97.83	94.10	95.38	99.25	98.39	98.92	91.29	90.12	86.61	94.34	89.97	83.72	99.53	98.98	98.36	89.79	89.63	77.20	90.53	88.51	77.68	95.42	90.77	85.72
36	12345678	97.78	95.34	96.23	97.00	94.18	94.70	99.54	97.74	97.91	94.68	90.96	85.13	94.70	90.32	84.15	99.42	98.38	98.53	91.45	88.21	78.63	91.37	88.32	76.37	93.69	91.15	85.47

CAC: Combined accuracy; PAC: Peak and lag accuracy; EAC: Entropy accuracy.

the MEEI database and SVM classifier, obtaining 99.99% with 45 features and 99.41% accuracy with only seven features. Moreover, Al-nasheri et al. [41] used the SVD, AVPD, and MEEI with the SVM classifier and used the MDVP parameters to detect the pathology, obtaining accuracies of 99.68%, 72.53%, and 88.21% for SVD, AVPD, and MEEI, respectively. In our study, the obtained accuracy in the case of using SVD is comparable (or slightly inferior) to the obtained accuracy in [41]; however, the three features used in [41] were databases dependent or tuned to a specific database; in our case, the features are database independent.

Table 10 shows the comparison between our method and the other methods used in different studies for detection by using MEEI, SVD, and AVPD. In our study, the highest obtained accuracies for detection were 99.69%, 92.79%, and 99.79% for MEEI, SVD, and AVPD, respectively, while the highest achieved accuracies for classification were 99.54%, 99.53%, and 96.02, respectively, using the combined feature.

VII. CONCLUSION

In this study, we evaluated the features (peak, lag, entropy, their combination) for different databases (MEEI, SVD, AVPD) and investigated the performance of these features on eight different filters with specific frequency bands in order to assess the effect of each band on the detection and classification processes. In this paper, the obtained accuracies from detection and classification varied for the same database as well as from one database to another with the same used feature. The best acquired accuracies for detection were 99.69%, 92.79%, and 99.79% for MEEI, SVD, and AVPD, respectively, while the highest achieved accuracies for classification were 99.54%, 99.53%, and 96.02, respectively, using the combined feature. Some of the frequency bands performed better compared with the others, but the performance of consecutive combined bands was better than that of the individual band. The best performance was in the bands of frequency range 1000∼8000 Hz. The obtained results for the cross-database remained almost the same in the three different databases with the three different features. In future work, we will investigate the variety of voice disorders and their severities.

APPENDIX A

See Table 11.

APPENDIX B

See Table 12.

REFERENCES

[1] G. Muhammad *et al.*, "Automatic voice pathology detection and classification using vocal tract area irregularity," *Biocybern. Biomed. Eng.*, vol. 36, no. 2, pp. 309–317, 2016.

[2] (2014). *National Institute on Deafness and Other Communication Disorders: Voice, Speech, and Language: Quick Statistics*, accessed on Mar. 2015. [Online]. Available: http://www.nidcd.nih.gov/health/statistics/vsl/Pages/stats.aspx

[3] *Research Chair of Voicing and Swallowing Disorders*, accessed on Oct. 2015. [Online]. Available: http://c.ksu.edu.sa/vas/en/vsb

[4] N. Roy, R. M. Merrill, S. Thibeault, R. A. Parsa, S. D. Gray, and E. M. Smith, "Prevalence of voice disorders in teachers and the general population," *J. Speech, Lang., Hearing Res.*, vol. 47, no. 2, pp. 281–293, Apr. 2004.

[5] K. H. Malki, "Voice disorders among Saudi teachers in Riyadh city," *Saudi J. Oto-Rhinolaryngol. Head Neck Surgery*, vol. 12, no. 2, pp. 189–199, 2010.

[6] B. Boyanov and S. Hadjitodorov, "Acoustic analysis of pathological voices. A voice analysis system for the screening of laryngeal diseases," *IEEE Eng. Med. Biol. Mag.*, vol. 16, no. 4, pp. 74–82, Jul. 1997.

[7] M. Chen, Y. Ma, J. Song, C. Lai, and B. Hu, "Smart clothing: Connecting human with clouds and big data for sustainable health monitoring," *Mobile Netw. Appl.*, vol. 21, no. 5, pp. 825–845, 2016.

[8] B. S. Atal, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," *J. Acoust. Soc. Amer.*, vol. 55, no. 6, pp. 1304–1312, 1974.

[9] M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, and C.-H. Youn, "Wearable 2.0: Enabling human-cloud integration in next generation healthcare systems," *IEEE Commun.*, vol. 55, no. 1, pp. 54–61, Jan. 2017.

[10] M. S. Hossain and G. Muhammad, "Cloud-assisted speech and face recognition framework for health monitoring," *Mobile Netw. Appl.*, vol. 20, no. 3, pp. 391–399, Feb. 2015.

[11] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[12] Z. Ali, M. Aslam, and A. M. M. Enriquez, "A speaker identification system using MFCC features with VQ technique," in *Proc. 3rd IEEE Int. Symp. Intell. Inf. Technol. Appl.*, Nov. 2009, pp. 115–119.

[13] M. S. Hossain, "Patient state recognition system for healthcare using speech and facial expressions," *J. Med. Syst.*, vol. 40, no. 12, pp. 272:1–272:8, Dec. 2016.

[14] J. I. Godino-Llorente, P. Gomez-Vilda, and M. Blanco-Velasco, "Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 10, pp. 1943–1953, Oct. 2006.

[15] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math. Statist.*, vol. 37, no. 6, pp. 1554–1563, 1966.

[16] S. Abe, *Support Vector Machines for Pattern Classification*. Berlin, Germany: Springer-Verlag, 2005.

[17] R. T. Ritchings, M. McGillion, and C. J. Moore, "Pathological voice quality assessment using artificial neural networks," *Med. Eng. Phys.*, vol. 24, no. 8, pp. 561–564, Sep. 2002.

[18] M. Chen, P. Zhou, and G. Fortino, "Emotion communication system," *IEEE Access*, vol. 5, pp. 326–337, 2017.

[19] M. D. O. Rosa, J. C. Pereira, and M. Grellet, "Adaptive estimation of residue signal for voice pathology diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 1, pp. 96–104, Jan. 2000.

[20] *Kay Elemetrics, Multi-Dimensional Voice Program (MDVP) [Computer Program]*, Lincoln Park, NJ, USA, 2012.

[21] M. K. Arjmandi, M. Pooyan, M. Mikaili, M. Vali, and A. Moqarehzadeh, "Identification of voice disorders using long-time features and support vector machine with different feature reduction methods," *J. Voice*, vol. 25, no. 6, pp. e275–e289, Nov. 2011.

[22] J. Wang and C. Jo, "Vocal folds disorder detection using pattern recognition methods," in *Proc. 29th Annu. Int. Conf. IEEE EMBS*, Lyon, France, Aug. 2007, pp. 3253–3256.

[23] T. Li, C. Jo, S.-G. Wang, B.-G. Yang, and H.-S. Kim, "Classification of pathological voice including severely noisy cases," in *Proc. 8th Int. Conf. Spoken Lang. Process.*, Jeju, South Korea, 2004, pp. 77–80.

[24] N. Sáenz-Lechón, J. I. Godino-Llorente, V. Osma-Ruiz, and P. Gómez-Vilda, "Methodological issues in the development of automatic systems for voice pathology detection," *Biomed. Signal Process. Control*, vol. 1, no. 2, pp. 120–128, Apr. 2006.

[25] M. A. Kiliç, H. Oguz, and M. A. Şafak, "Comparison of results in two acoustic analysis programs: PRAAT and MDVP," *Turkish J. Med. Sci.*, vol. 41, no. 5, pp. 835–841, 2011.

[26] G. Muhammad and M. Melhem, "Pathological voice detection and binary classification using MPEG-7 audio features," *Biomed. Signal Process. Controls*, vol. 11, pp. 1–9, May 2014.

[27] G. Muhammad, T. A. Mesallam, K. H. Malki, M. Farahat, A. Mahmood, and M. Alsulaiman, "Multidirectional regression (MDR)-based features for automatic voice disorder detection," *J. Voice*, vol. 26, no. 6, pp. 817.e19–817.e27, 2012.

[28] A. Al-nasheri, Z. Ali, G. Muhammad, and M. Alsulaiman, "Voice pathology detection using auto-correlation of different filters bank," in *Proc. 11th ACS/IEEE Int. Conf. Comput. Syst. Appl.*, Doha, Qatar, Nov. 2014, pp. 50–55.

[29] V. von Leden, Hans, P. Moore, and R. Timcke, "Laryngeal vibrations: Measurements of the glottic wave—Part III. The pathologic larynx," *AMA Arch. Otolaryngol.*, vol. 71, no. 1, pp. 16–35, 1960.

[30] P. Lieberman, "Perturbations in vocal pitch," *J. Acoust. Soc. Amer.*, vol. 33, no. 5, pp. 597–603, 1961.

[31] D. Mansour and B. H. Juang, "The short-time modified coherence representation and noisy speech recognition," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 37, no. 6, pp. 795–804, Jun. 1989.

[32] *Disordered Voice Database, Version 1.03 (CD-ROM), MEEI, Voice and Speech Lab*, Kay Elemetrics Corp., Boston, MA, USA, Oct. 1994.

[33] W. J. Barry and M. Pützer. Saarbrucken voice database. Institute of Phonetics, University of Saarland. Accessed: Mar. 10, 2017. [Online]. Available: http://www.stimmdatenbank.coli.uni-saarland.de/

[34] D. Martínez, E. Lleida, A. Ortega, A. Miguel, and J. Villalba, "Voice pathology detection on the Saarbrücken voice database with calibration and fusion of scores using multifocal toolkit," in *Advances in Speech and Language Technologies for Iberian Languages*. Berlin, Germany: Springer, 2012, pp. 99–109.

[35] M. Markaki and Y. Stylianou, "Voice pathology detection and discrimination based on modulation spectral features," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 7, pp. 1938–1948, Sep. 2011.

[36] R. Fraile, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, and J. M. Gutiérrez-Arriola, "Characterization of dysphonic voices by means of a filterbank-based spectral analysis: Sustained vowels and running speech," *J. Voice*, vol. 27, no. 1, pp. 11–23, 2013.

[37] G. Pouchoulin, C. Fredouille, J.-F. Bonastre, A. Ghio, and J. Revis, "Characterization of the pathological voices (dysphonia) in the frequency space," in *Proc. Int. Congr. Phonetic Sciences (ICPhS)*, Aug. 2007, pp. 1993–1996.

[38] R. D. Kent and Y. Kim, "Acoustic analysis of speech," in *The Handbook of Clinical Linguistics*, M. J. Ball, M. R. Perkins, N. Müller, and S. Howard, Eds. Oxford, U.K.: Blackwell Publishing Ltd., 2008, ch. 22, pp. 364–365, doi: 10.1002/9781444301007.

[39] J.-W. Lee, H. G. Kang, J.-Y. Choi, and Y. I. Son, "An investigation of vocal tract characteristics for acoustic discrimination of pathological voices," *BioMed Res. Int.*, vol. 2013, Art. no. 758731, Sep. 2013, doi: 10.1155/2013/758731.

[40] G. Muhammad, M. Alsulaiman, A. Mahmood, and Z. Ali, "Automatic voice disorder classification using vowel formants," in *Proc. IEEE Int. Conf. Multimedia Expo Workshop ((ICME))*, Barcelona, Spain, Jul. 2011, pp. 1–6.

[41] A. Al-Nasheri, Z. Ali, G. Muhammad, and M. Alsulaiman, "An investigation of MDVP parameters for voice pathology detection on three different databases," in *Proc. 16th Annu. Conf. Int. Speech Commun. Assoc. (INTERSPEECH)*, Dresden, Germany, Sep. 2015, pp. 2952–2956.

AHMED AL-NASHERI received the master's degree in computer science with specialization in programming language from New Mexico State University, USA, in 2003, and the Ph.D. degree from the Computer Engineering Department, King Saud University, Riyadh, Saudi Arabia. He is associated with the Department of Computer Science, Sana'a Community College. His research area is speech processing and voice pathology.

GHULAM MUHAMMAD (M'07) received the Ph.D. degree in electrical and computer engineering with specialization in speech recognition from the Toyohashi University of Technology, Japan, in 2006. He is currently a Professor with the Computer Engineering Department, King Saud University, Riyadh, Saudi Arabia. He was a recipient of the Japan Society for Promotion of Science Post-Doctoral Fellowship. He has authored and co-authored around 150 publications, including refereed IEEE/ACM/Springer/Elsevier journals and conference papers. He holds a U.S. patent on environment. His research interest includes speech processing, image processing, smart health, and multimedia signal processing. He is a member of the ACM.

MANSOUR ALSULAIMAN received the Ph.D. degree from Iowa State University, USA, in 1987. Since 1988, he has been associated with the Department of Computer Engineering, College of Computer and Information Sciences. He is currently a Professor with the Computer Engineering Department, King Saud University, Riyadh, Saudi Arabia. His research areas include automatic speech/speaker recognition, automatic voice pathology assessment systems, computer-aided pronunciation training system, and robotics.

ZULFIQAR ALI received the M.S. and M.Sc. degrees in computer science with the specialization in system engineering from the University of Engineering and Technology, Lahore, and the M.Sc. degree in computational mathematics from the University of the Punjab, Lahore, and the Ph.D. degree in electrical and electronic engineering from the Center for Intelligent Signal and Imaging Research, Universiti Teknologi Petronas, Malaysia. He is currently a Researcher with the Digital Speech Processing Group, Department of Computer Engineering, King Saud University, Saudi Arabia. His research and teaching career spans over 16 years. He has over 45 publications in journals and conferences of international repute and has a very active research profile. He has the leading roles in many funded projects. His research interests include speech and language processing, cloud and multimedia for healthcare, privacy and security, watermarking, steganography, and multimedia forensics.

KHALID H. MALKI received the MBBS degree from King Faisal University, Saudi Arabia, and the Ph.D. degree from Ein Shams University, Egypt, in 2004. He is a Professor of Phoniatrics (Communication and Swallowing Disorders), ENT Department, King Saud University, Riyadh, Saudi Arabia. His research areas include voice, swallowing, language, and speech disorders.

TAMER A. MESALLAM is a Professor of Phoniatrics with the ENT Department, College of Medicine, King Saud University, Saudi Arabia. He is the Deputy Chairman of Communication and Swallowing Disorders Unit and the Supervisor of Laryngopharyngeal Reflux Clinics with King Abdulziz University Hospital. He has spent two years as a Research Fellow with the Voice Center, Cincinnati Children's Hospital, OH, USA. He is currently the Director of the ENT Research Unit with King Abdulaziz University Hospital and the Research Director of the Research Chair of Voice, Swallowing, and Communication Disorders. He also has been currently appointed as the Head of ENT Committee for Clinical Practice Guidelines with the College of Medicine, King Saud University. He has several publications in the field of voice and communication disorders.

MOHAMED FARAHAT IBRAHIM received the bachelor's degree in medicine and general surgery, the M.Sc. degree, and the M.D. degree in phoniatrics from the College of Medicine, Ain Shams University, Cairo, Egypt. From 2008 to 2016, he was the Chairman of the Communication and Swallowing Disorders Unitwith King AbulAziz University Hospital, King Saud University, where he was the Supervisor of Swallowing Disorders Clinics with King Khalid University Hospital, College of Medicine. In addition to academic and clinical assignments, he was the coordinator of the ORL course for the fourth year medical students, College of Medicine, King Saud University, for three years. He spent four years in the ENT Team for Canadian Accreditation, ENT Department, King Saud University. He was nominated as the International Coordinator of World Swallowing Day from the Union of European Phoniatricians. From the research point of view, he is heading the King Saud Research Group of Voice, Swallowing, and Communication Disorders and also a member in a grant funded by the National Plan of Science and Technology. He is currently a Professor and Consultant Phoniatrician (Voice, speech, language, and swallowing disorders). He has authored or co-authored over 19 articles with his colleagues in the field of voice, swallowing, and communication disorders. He has been a member of the Research Chair of Voice, Swallowing and Communication Disorders since 2009.