An Intelligent Healthcare System for Detection and Classification to Discriminate Vocal Fold Disorders

[1]Zulfiqar Ali, [2]M. Shamim Hossain, [1,3]Ghulam Muhammad and [4]Arun Kumar Sangaiah
[1]Digital Speech Processing Group, College of Computer and Information Sciences
King Saud University, Riyadh 11543, Saudi Arabia
[2]Department of Software Engineering, College of Computer and Information Sciences
King Saud University, Riyadh 11543, Saudi Arabia
[3]Department of Computer Engineering, College of Computer and Information Sciences
King Saud University, Riyadh 11543, Saudi Arabia
[4]School of Computing Science and Engineering, VIT University, Vellore-632014, India

Corresponding Author: mshossain@ksu.edu.sa

ABSTRACT

The growing population of senior citizens around the world will appear as a big challenge in the future and they will engage a significant portion of the healthcare facilities. Therefore, it is necessary to develop intelligent healthcare systems so that they can be deployed in smart homes and cities for remote diagnosis. To overcome the problem, an intelligent healthcare system is proposed in this study. The proposed intelligent system is based on the human auditory mechanism and capable of detection and classification of various types of the vocal fold disorders. In the proposed system, critical bandwidth phenomena by using the bandpass filters spaced over Bark scale is implemented to simulate the human auditory mechanism. Therefore, the system acts like an expert clinician who can evaluate the voice of a patient by auditory perception. The experimental results show that the proposed system can detect the pathology with an accuracy of 99.72%. Moreover, the classification accuracy for vocal fold polyp, keratosis, vocal fold paralysis, vocal fold nodules, and adductor spasmodic dysphonia is 97.54%, 99.08%, 96.75%, 98.65%, 95.83%, and 95.83%, respectively. In addition, an experiment for paralysis versus all other disorders is also conducted, and an accuracy of 99.13% is achieved. The results show that the proposed system is accurate and reliable in vocal fold disorder assessment and can be deployed successfully for remote diagnosis. Moreover, the performance of the proposed system is better as compared to existing disorder assessment systems.

Keywords HealthcareVocal fold disordersBinary classificationCritical bandsAuditory perception

I. INTRODUCTION

Due to rapid growth in information and communication technologies, the building of smart homes and cities becomes a reality. Smart homes and cities take the home and living experience to the next level. One of the major reasons for the development of smart homes and cities is to provide the efficient and cost-effective healthcare facilities. According to the American Association of Retired Persons [1], 85% of senior citizens want to stay at home for the treatment as long as the facilities are available. Of concern is that a large population around the world is aged 60 years or above. In the report of the United Nations on world population aging which was published in 2015 [2], it is mentioned that around 900 million people around the world are 60 years of age or above. This population will rise to 1402 million until 2030. Such a large population will occupy a significant portion of the health facilities in the hospitals. This situation can be avoided by building smart homes and cities, where automatic

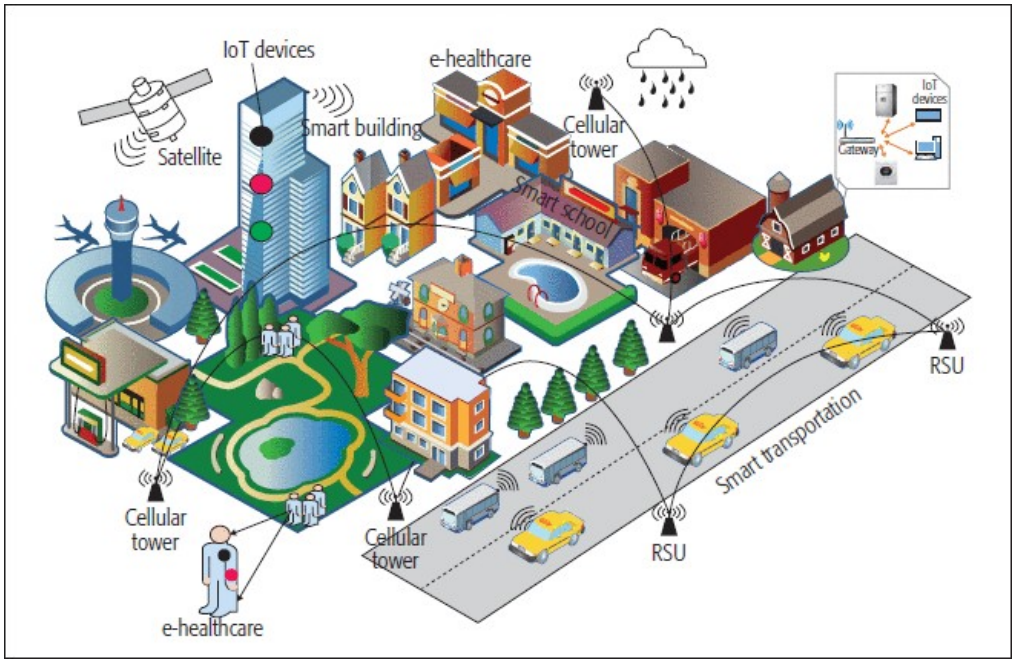


Figure 1. An illustration of an IoT-based smart city [3].

diagnosis systems will be a critical component [4-8]. The automatic healthcare systems recieve the data through the Internet of Things (IoT) and transmit it for the evaluation.

In smart cities, the IoT gathers a huge quantity of data and it can be processed by using automatic assessment systems [5]. A high-level illustration of an IoT-based smart city is depicted in Fig. 1. However, the increasing use of wireless transmission of health-related data raises the concern of data protection and authenticity. The medical data of an individual may be secured such that unauthorized access to the data could be denied. Only authorized healthcare staff may access the data to ensure the privacy of an individual's identity. A framework for a privacy-protected healthcare is shown in Fig. 2.

The main goal of this study is to develop a detection and classification system for vocal fold disorders which can be deployed in smart homes and cities for automatic diagnosis. A voice disorder affects the vocal folds and makes the vibration of the vocal folds abnormal. The characteristics of various voice disorders such as vocal fold nodules, cysts, and paralysis are presented in [9]. Due to irregular vibrations, the vocal folds exhibit incomplete closure or tight closure, which makes the voice breathy, weaker, strained, and harsh. The abnormal behavior of the vocal folds disturbs voice patterns and therefore the speech signal of a disordered person becomes more transient and noisy compared to that of a normal person [10]. A large number of populations around the world suffer from different kinds of voice disorders. According to the National Institute on Deafness and Other Communication Disorders, approximately 17.9 million people suffer from voice problems [11]. Around 700 cases of voice complications per year are observed in Riyadh, Saudi Arabia. More than 15% of the people visiting King Abdul Aziz University complain about voice problems [12]. Various types of voice disorders are described as follows.

Vocal fold polyps are fluid-filled lesions that appear on the free edge of the vocal folds, and the main reason for their occurrence is the abuse of the voice. Polyps resemble a blister; they are reddish in color. Polyps are associated with frequent breaks in singers, earlier vocal fatigue, and worsening dysphonia [13]. Several factors can contribute to the formation of vocal fold polyps, such as allergies, nicotine, and voice trauma [14]. Vocal fold polyps usually occur in adult men who use their voices excessively; these patients also have a high risk of vocal fold nodules and cysts [15].

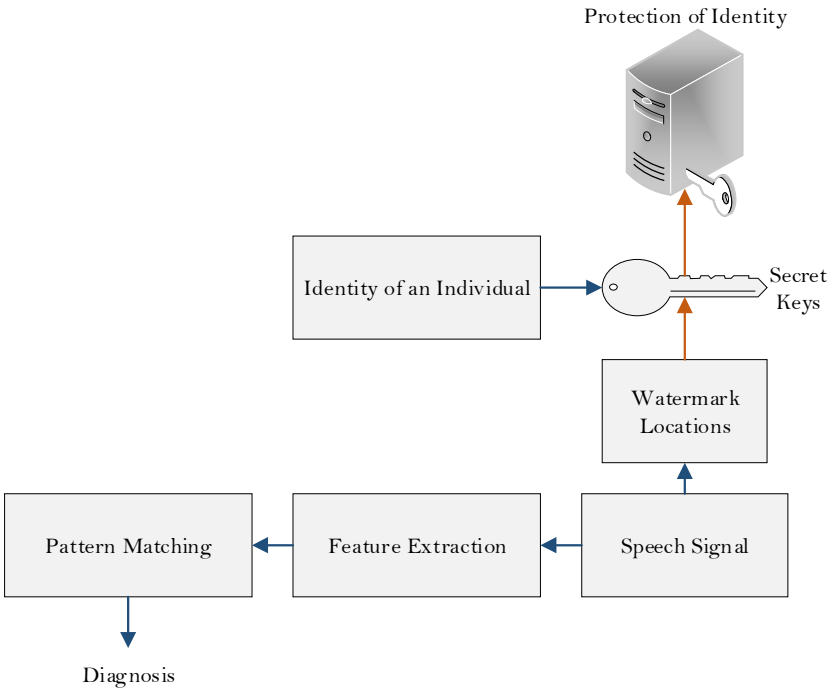


Figure 2. A framework for the privacy protected healthcare system.

Keratosis appears due to the presence of abnormal cells (white plaques) on the vocal folds [16]. This lesion is pre-cancerous but can turn into cancer in the case of negligence. Keratosis disturbs the normal vibration of the vocal folds and causes hoarseness. The main reason for the development of keratosis is smoking, the excessive use of the voice, and environmental pollutants. Gastroesophageal reflux disease may also be a reason to produce abnormal changes in cells [17]. Keratosis may be unilateral or bilateral and usually symmetric in nature. This lesion has more tendency to prevail in men than in women.

Vocal fold paralysis occurs due to the malfunctioning of one or both vocal folds when they open and close improperly. Unilateral vocal fold paralysis is a common disorder; however, bilateral vocal fold paralysis is rare and life-threatening. One of the main reasons for vocal fold paralysis is an injury to the recurrent laryngeal nerve [18]. This nerve controls the motion of the vocal folds. Vocal fold paralysis can also occur due to injury to the chest, neck, or head; thyroid or lung cancer; and tumors on the chest, neck, or skull base.

Vocal fold nodules [19] occur as bilateral symmetric swelling located at the junction of the anterior and mid-third part of the vocal folds. This is the point of the maximal shearing and collision forces between the folds. Usually, nodules affect adult women and male adolescents and can vary in color, size, and symmetry. Common symptoms of nodules are hoarseness, breathiness, and vocal breaks. Vocal fold nodules are common among persons who use their voices chronically such as teachers, stock traders, and singers.

Spasmodic dysphonia is a neurological disorder that affects the muscle of the human voice box (larynx) [20]. In spasmodic dysphonia, when air pressure from the lungs vibrates the vocal folds, a voluntary movement inside the muscles of the vocal folds (called spasms) is produced, which affects the vibration of the vocal folds. *Adductor spasmodic dysphonia* is a type of spasmodic dysphonia in which spasms cause the vocal folds to slam together and stiffen [21]. Vocal folds experience difficulty in vibration due to these spasms; therefore, a person feels a problem in voice production. Usually, words are cut off or harder to start due to these muscles. Adductor spasmodic dysphonia makes the voice of a person pressed, strangled, strained, and full of effort.

Screening a disorder in the initial stage may reduce health complications, and a person can visit a respective medical specialist to cure voice complications. Most voice disorders occur due to the excessive use of the voice. Therefore, people involved in the professions of teaching [22], singing, and stock markets have a high risk of voice disorders. In the USA, the prevalence of voice disorders in teachers is 57.7% during their lifetime, and for other professions, it is 28.8% [23]. The intelligent systems for detection and classification of vocal fold disorders can be developed for the early screening so that the patients can avoid the sever circumstance caused by negligence and delay in diagnosis.

Most of the existing intelligent systems are developed to detect the vocal fold disorders [24-28]. The system developed in [24] is implemented by using shimmer, jitter and signal-to-noise ratio. The obtained accuracy of the system is 80.3%, which is not good. The reason is the dependency of the features on fundamental frequency (F0), and the estimation of F0 is itself a challenging task because disordered speech signals are aperiodic in nature [29-31]. The detection system developed in [25] also based on F0-dependent features, and the obtained accuracy is again very low, i.e., 70%. Another detection system based on F0-dependent features and cepstral coefficients is developed in [26]. The achieved accuracy of the system is 91.32%, which is comparatively better. A disorder detection system based on multiresolution analysis of the normal and disordered signal is developed in [27]. Various frequency regions of speech signals are investigated by using discrete wavelet transformation and fractal dimensions to determine the frequencies that can contribute significantly to the detection of disorders. The highest obtained accuracy is 92.45%. Moreover, a software for the screening of dysphonic patients is developed in [28]. The developed software differentiates between normal and disordered subjects by analyzing the recurrence plots of the speech signals with local binary pattern (LBP) operator. The maximum obtained detection accuracy of the software is 97.73%. These existing systems only determine the presence of disorders but cannot differentiate among various types of the vocal fold disorders.

In [32], a system for the classification of various types of disorders is developed by using Mel-frequency cepstral coefficients (MFCC). Three types of vocal fold disorders vocal fold nodules, vocal fold edema, and unilateral paralysis are considered to develop the system. The obtained accuracy is 66% for the Gaussian mixture model (GMM) and 69% for the support vector machine (SVM). The results show that the performance of the system is not satisfactory. Moreover, some systems exist in the literature that can perform both types of tasks, disorder detection as well classification [33, 34]. The diagnosis of disorders in such systems based on the decision of automatic classifier but they do not provide any visual indication for the presence of voice disorders. The user of an automatic system can diagnose the disorder more accurately and reliably if it is supported by a clear visual indication.

In this study, an intelligent vocal fold disorder assessment system is proposed which can perform both types of tasks, disorder detection and classification. In addition, the proposed system provides a clear visual indication for the presence of vocal fold disorders. Similar to the human auditory system, the developed system detects and classifies disordered and normal samples by using the principle of the human auditory system. The structure of the human ear is such that it can differentiate frequencies of speech or voice signals. The inner ear has a spiral cochlea, which is surrounded by the basilar membrane. The higher frequencies are observed with the excitation at the basal turn, while the lower frequencies are observed towards the apex of the cochlea [35]. Each region of the basilar membrane acts like a band-pass filter, and the width of the filter follows the critical bandwidth of the human hearing perception. In this study, the phenomena of critical bandwidth is implemented by using the bank of bandpass filters spaced over Bark scale. As compared to the Mel scale, the Bark scale reduces the linearity. The detection and classification results of the proposed systems are good and better than the existing systems.

The restof the paper is structured as follows. Section 2 describes the proposed automatic detection and classification system. The experiments for disorder detection and classification are provided in Section 3. The analysis of the proposed method is presented in Section 4. Finally, Section 5 draws some conclusions.

II. MATERIALS AND METHODS

Speech features to perform disorder detection and classification are extracted through acoustic analysis of speech signals in this study. Then, these extracted features are used with GMM to develop the proposed system. A publicly available voice disorder database is considered to evaluate the proposed system.

A. Material

To evaluate the performance of the proposed intelligent system for the detection and classification of voice disorders, the Massachusetts Eye & Ear Infirmary (MEEI) voice disorder database is considered. The database is recorded at the MEEI voice and speech laboratory [36] and has been used in the number of studies [33, 37-41]. The MEEI database contains 710 voice samples. It contains 53 samples of normal subjects and 657 samples of pathological subjects suffering from more than 100 types of voice disorders. The pathological data are recorded by patients suffering from a variety of voice disorders. Patients were asked to produce a sustained vowel /ah/, which was recorded at a 25 kHz sampling frequency with a bit rate of 16-bit. Some of the samples were recorded at 50 kHz; therefore, these samples are down-sampled to 25 kHz before performing the acoustic analysis. The normal data were recorded by the people who do not have any history of voice problems. The duration of the sustained vowel for a patient is one second, and that is three seconds for a normal subject. The possible reason for the shorter duration of the signal is that the patients cannot hold the breath for a long duration due to the pain he or she is suffering from. All samples of the MEEI database are used in the acoustic analysis to extract the speech features.

B. Acoustic Analysis of Speech Signals

Speech varies quickly over time, and the analysis becomes difficult due to its dynamic nature. Therefore, it is necessary to divide the whole speech signal into short frames for accurate decision. In this study, each speech signal is partitioned into frames of 256 samples. Partitioning of the speech signal into short frames makes the signal stationary. Moreover, each current frame has 50% overlap with the previous frame to avoid the loss of information at the endpoints. In addition, endpoints of frames cause spectral leakage during the implementation of Fourier's transformation (FT). This problem can be avoided by tapering the ends of each frame to zero, and it can be done by multiplying the frames with the hamming window *h(n)* [42]. Furthermore, application of the hamming window ensures the continuity between frames of a signal. The window is given by Eq. (1), where *N* represents the number of samples in each frame.

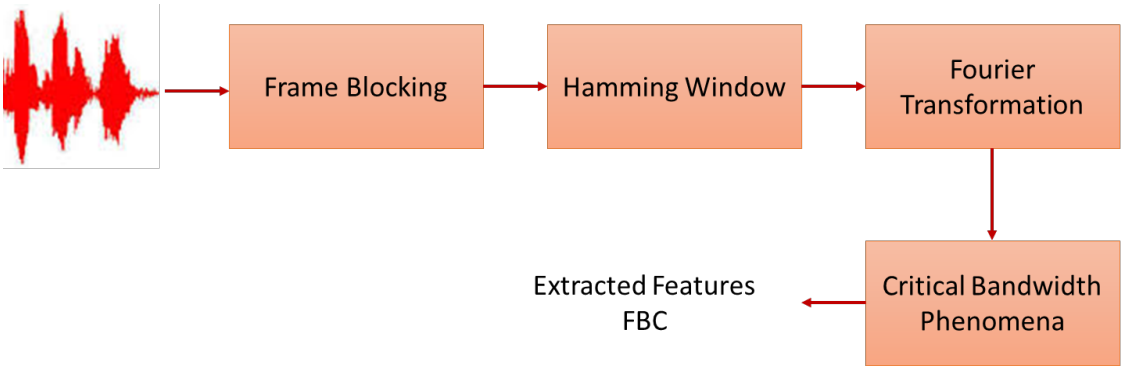
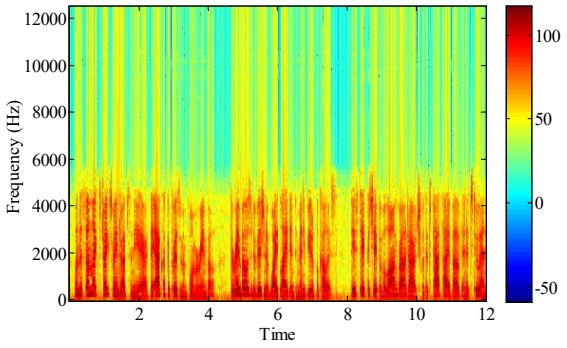


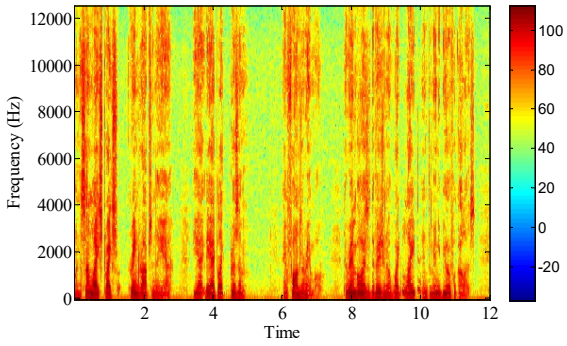
Figure. 3. The steps of acoustic analysis for extraction of the feature FCB.

$$h(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) \quad \text{where} \quad 0 \le n \le N-1 \qquad (1)$$

The block diagram to perform the acoustic analysis for extraction of features is shown in Fig. 3. The extracted features are based on the phenomena of critical bandwidths, and they are referred as FCB. The next step in the calculation of FCB, after hamming window, is the implementation of FT to transform the speech signal from the time domain to the frequency domain. FT provides the information of energy in each frequency element, and the output after applying FT on a speech signal is known as a spectrum. The spectrum of a normal and disordered subject is depicted in Fig. 4. It can be noticed in Fig. 4(a) that the energy is contained in the lower frequency components of the spectrum for normal persons. On the other hand, the energy in the spectrum of the disordered patients is spread over all frequency components as shown in Fig. 4(b).



(a)



(b)

Figure 4. Spectrums of speech signals (a) Normal (b) Disordered.

The human ear does not perceive frequencies in a linear way. The ear has more capability to differentiate between lower frequencies than the higher frequencies. Therefore, each frame of the spectrum is passed through a band-pass filter to simulate the human hearing system. The bandwidth of each band-pass filter is referred as a critical bandwidth. In this study, 29 band-pass filters are used. The filters are determined by applying Bark scale proposed by Zwicker [43]. According to this scale, the bandwidth is linear up to 500 Hz, and then, increased by 20% of the center frequency of a band above 500 Hz. The scale is given by Eq. (2), where z represents the frequency in Hz and B stands for corresponding frequency in Bark scale.

$$B = 13\arctan(0.00076z) + 3.5\arctan\left(\frac{z}{7500}\right)^2 \qquad (2)$$

After applying the critical band-pass filter, the required features FCB are obtained. Then, FCB is given to a GMM based classifier for automatic detection and classification of voice disorders.

C. Development of Automatic Classifier

The features FCB are computed from all normal and disordered speech signals one by one. For each signal, the dimension of the computed features FCB is R x S. The parameters R and S represent the number of the frames in a speech signal and bandpass filters, respectively. For instance, if the duration of a speech signal is one second and the sampling frequency is 25 kHz, then the dimension of FCB will be 194x24 and interpretation of such a multidimensional features is impossible by the human mind. Therefore, a pattern recognition phase becomes necessary to find the trend in the computed features. In this study, we have two major tasks. The first task is a differentiation between normal and disordered signals, and the second task is a classification of different types of voice disorders. In both types of tasks, the most crucial thing is the extraction of patterns from features that is achieved by developing an automatic classifier.

A state-of-the-art clustering technique of pattern recognition is implemented to develop the classifier. This clustering technique is GMM [44], and it has been used in many scientific disciplines [45-47]. GMM is used to generate acoustic models for speech signals of various classes by using the computed features FCB. GMM is preferred over other clustering techniques such as k-means algorithm [48]. K-means is based on hard assignments in which each data point belongs to one cluster, while GMM assumes that a data point can belong to different clusters where the probability for each cluster is different. However, to develop the acoustic model in this study, k-means is used to initialize the parameters of GMM. Moreover, these parameters are estimated and adjusted by using the expectation-maximization (EM) algorithm [49]. The tuned parameters provide acoustic models with maximum log-likelihood (LL).

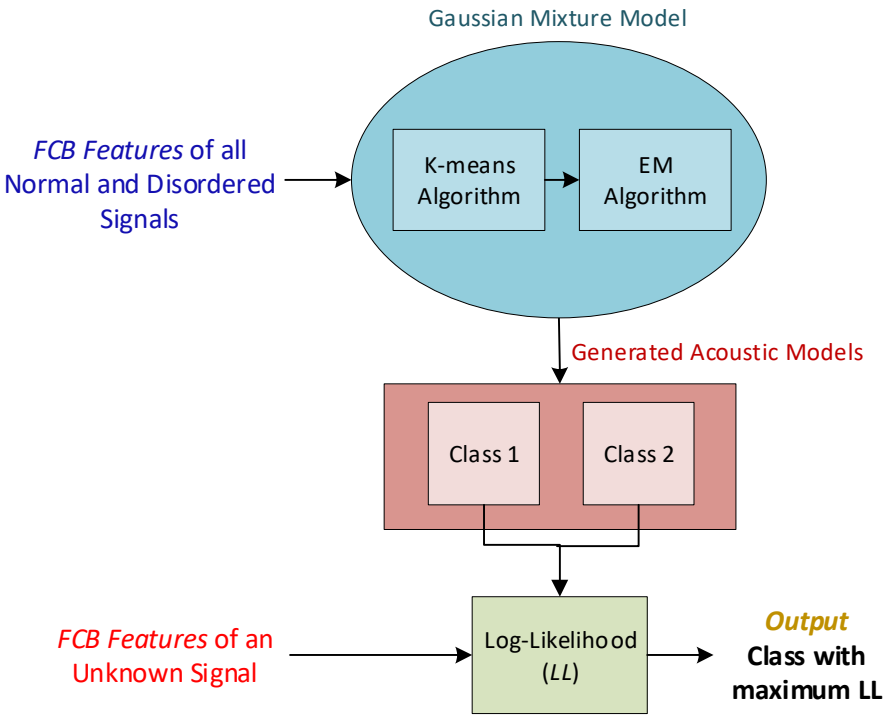


Figure 5. Block diagram for the proposed detection and classification system.

The developed automatic classifier, shown in Fig. 5, is used to generate the proposed detection and classification system for voice disorders. The proposed system consists of two major steps. The first step is the training of the system, and the second step is testing of the system. In the training step, the proposed system extracts the FCB

features from all normal and disordered signals. Then, the FCB features are given to GMM for generation of acoustic models for each class. During the testing phase, these generated models are compared with the FCB features of an unknown signal and LL of the test signal is computed with each model. The model having maximum LL will be the class of the test utterance. In this way, the proposed system determines the types of voice disorders.

III. EXPERIMENTAL SETUP AND RESULTS

The proposed detection and classification system is evaluated by conducting many different types of experiments. To obtain the reliable results, the proposed system is tested with every single voice sample of the MEEI database. For this, we have used a three folds cross-validation approach. In this approach, the MEEI database is partitioned into three disjoint subsets. The system is tested with one of the subsets, while the remaining two are used for the training of the system. To report the results of the proposed system, various measures are considered. These measures are sensitivity, specificity, accuracy, and area under the receiver operating characteristic (ROC) curve. The sensitivity is a ratio between correctly identified disordered signals and the total number of voice disordered signals. The specificity is a ratio between correctly identified normal speech signal and the total number of normal speech signals. The accuracy of the system describes that how many speech signals are correctly identified from the total number of normal and disordered signals. These performance measures are computed by the relations given in Eq. (3)-(5). Moreover, ROC curve graphically shows the performance of a binary classifier. If the area under the ROC curve (AUC) is close to one, it means the results of the system are reliable.

$$Sensitivity = \frac{True\ Disordered}{True\ Disordered + False\ Healthy} \times 100 \quad (3)$$

$$Specificity = \frac{True\ Healthy}{True\ Healthy + False\ Disordered} \times 100 \quad (4)$$

$$Accuracy = \frac{True\ Disordered + True\ Healthy}{Total\ Numer\ of\ Samples} \times 100 \quad (5)$$

where *True Disordered* represents that a disordered signal is also identified as a disordered signal by the system, *False Healthy* means that a disordered signal is identified as a healthy signal by the system, *True Healthy* denotes that a healthy signal is also identified as a healthy signal by the system, *False Disordered* means a healthy signal detected as a disordered signal by the system, and *Total Number of Samples* stands for the total number of healthy and disordered samples.

Another set of features is extracted from the FCB by computing their first order derivatives to observe the variation in frequency over time. These features are referred as delta features of the FCB and calculated by using Eq. (6). The derivative is calculated with the regression window of length W. Moreover, $C_{T,F}$ represents the component of the FCB at the T^{th} frame and the F^{th} band-pass filter in Eq. (6).

$$\Delta_T = \frac{\sum_{K=1}^{W} K\left(C_{T-1,F} - C_{T+1,F}\right)}{2\sum_{K=1}^{W} K^2} \quad (6)$$

During extraction of FCB features, the phenomena of critical bandwidth to simulate the human hearing system is applied by using 29 band-pass filters, i.e., F=29. Therefore, the number of coefficients in the FCB for each frame is 29. The linear regression with W=5 is computed for the FCB by using Eq. (6). The regression provides 29 more coefficients. The combination of FCB and its derivative is represented by the FCBD. The number of coefficients in

the FCBD is 58. The other parameters of the systems are: the duration of each frame is 256, the current frame contains the 50% of the samples from the previous frame, and the numbers of points in hamming window and FT are 256.

A. Detection Results for Voice Disorders

For voice pathology detection, the experiments are conducted by using both types of features, FCB and FCBD. The experimental results for the detection of voice disorders with the FCB are provided in Table 1. These results are obtained with various numbers of Gaussian mixtures, i.e., 8, 16, 32 and 64. The maximum accuracy is obtained with 32, and this is 99.45%. Moreover, detection experiments with the FCBD are also performed by using the same numbers of Gaussian mixtures, and the results are presented in Table 2. The highest obtained accuracy with the FCBD is 99.72% with 64 mixtures. The obtained accuracy with 2 and 4 Gaussian mixtures was not good as compared to 8 mixtures. Therefore, these results are not listed in Table 1 and 2. In addition, Gaussian more than 64 do not provide a significant improvement in the accuracy. Therefore, these results are also not provided in Table 1 and 2.

A comparison between the accuracies of the FCB and FCBD for the different number of Gaussian mixtures is depicted in Fig. 6. The best detection rate is almost the same for the FCB and FCBD. However, the result of the FCB is better than that of the FCBD as it is obtained with fewer features and Gaussian mixtures. Fewer coefficients and Gaussians mixtures mean fewer computations and less running time. It can be observed from Table 1 and Table 2 that the standard deviation (*STD)* among the accuracy of the three folds decreased as the number of mixtures increased. This means the results obtained with higher numbers of Gaussian mixtures are more reliable.

Table 1. Voice disorder detection results for the FCB

Number of Gaussians	FCB			
	Sensitivity	*Specificity*	*Accuracy ± STD*	*AUC*
8	96.4	91.7	95.95 ± 1.2	0.9815
16	97.3	94.5	97.26 ± 1.0	0.9863
32	99.6	98.3	99.64 ± 0.8	0.9925
64	98.2	95.6	97.91 ± 0.7	0.9974

Table 2. Voice disorder detection results for the FCBD

Number of Gaussians	FCBD			
	Sensitivity	*Specificity*	*Accuracy ± STD*	*AUC*
8	93.6	90.1	93.56 ± 1.1	0.9579
16	94.9	91.2	94.83 ± 0.9	0.9669
32	95.9	93.4	95.85 ± 0.7	0.9799
64	99.6	98.9	99.72 ± 0.5	1

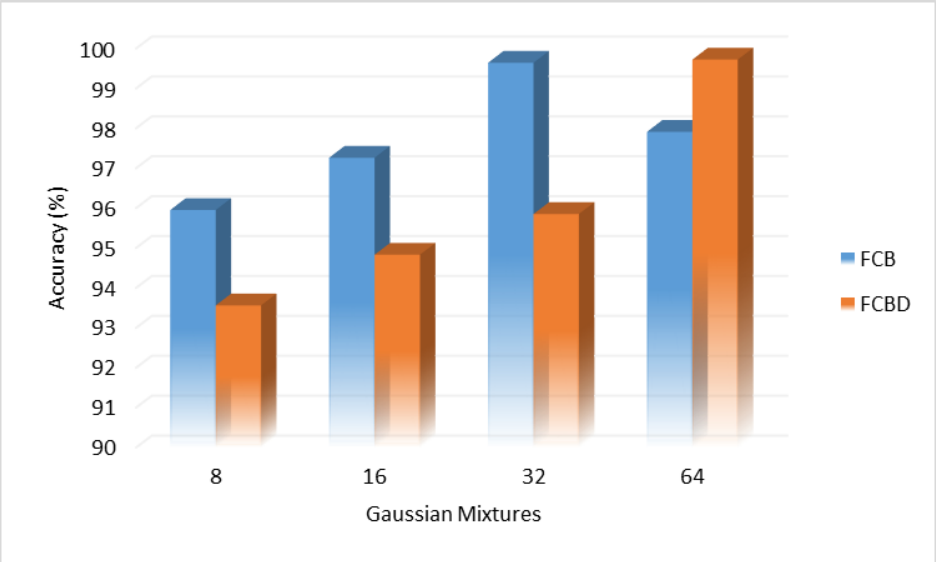


Figure 6. Comparison between accuracies of the FCB and FCBD.

The results of the disorder detection of the proposed system are very good. The overall highest accuracy 99.72% has an STD equal to 0.5 and AUC equivalent to 1. The small STD of 0.5 among accuracies of different folds suggests that the accuracy of the system does not change by changing the training and testing samples. Hence, the system is robust against the training and testing samples. In addition, the maximum AUC of 1 indicates that the system is stable in the detection of disorders and can be used reliably for remote diagnosis of the vocal fold disorders.

B. Classification Results for Voice Disorders

We have used the same setup for the classification of voice disorders as was used in [33, 34]. By using the same experimental setup, we will be able to compare the results of our proposed system with the existing systems of [33, 34]. The list of the experiments for the classification of voice disorders is given in Table 3. In the experiments, only those disorders that have at least 20 voice samples in the MEEI disorder database are considered. Those disorders are vocal fold polyp, keratosis, vocal fold paralysis, vocal fold nodules, and adductor spasmodic dysphonia. The number of samples for each experiment are mentioned in Table 3. It can be noted that the number of samples for polyp in experiment Ex1 is 20, while in Ex2 it is 17. The reason is that in Ex2, three patients suffer from polyp and keratosis at the same time. Therefore, these samples are excluded because they cannot be used for the classification of polyp and keratosis. The names of all voice disorder files used in this study are mentioned in Appendix A of [33].

In each experiment, a binary classification is performed. For example, in Ex1, the classification is performed between polyp and adductor. In all experiments, the first disorder is considered a positive class and the second disorder is considered a negative class. For Ex1, polyp is a positive class and adductor is a negative class. All classification experiments are carried out by using the FCB and FCBD. The experiments for the classification of polyp with adductor, keratosis, and nodules are conducted with 4, 8, and 16 mixtures. However, only the best results are reported in Table 4. The best obtained accuracy with the FCB for polyp vs. adductor is 96.83%, polyp vs. keratosis is 98.58%, and polyp vs. nodules is 96.14%. Moreover, the highest classification accuracy with the FCBD for polyp vs. adductor is 97.54%, polyp vs. keratosis is 99.08%, and polyp vs. nodules is 96.75%.

Table 3. The list of experiments for the classification of disorders taken from [33, 34]

Experiments	Classification	Number of Files
Ex1	Polyp and Adductor	Polyp: 20 and Adductor: 22
Ex2	Polyp and Keratosis	Polyp: 17 and Keratosis: 23
Ex3	Polyp and Nodules	Polyp: 19 and Nodules: 19
Ex4	Adductor and Nodules	Adductor: 22 and Nodules: 20
Ex5	Adductor and Keratosis	Adductor: 22 and Keratosis: 26
Ex6	Keratosis and Nodules	Keratosis: 26 and Nodules: 20
Ex7	Paralysis and Others	Paralysis: 71 and Others: 71

Table 4. Classification results of polyp with adductor, keratosis, and nodules

Performance Measures	Polyp and Adductor		Polyp and Keratosis		Polyp and Nodules	
	FCB	*FCBD*	*FCB*	*FCBD*	*FCB*	*FCBD*
Sensitivity	96.67	97.67	98.04	98.63	96.49	97.02
Specificity	96.97	97.42	98.99	99.42	95.79	96.49
Accuracy	96.83 ± 1.0	97.54 ± 0.9	98.58 ± 0.8	99.08 ± 1.1	96.14 ± 0.7	96.75 ± 1.2
AUC	0.9808	0.9921	0.9963	1	0.9758	0.9995

Table 5. Classification results of adductor with nodules and keratosis

Performance Measures	Adductor and Nodules		Adductor and Keratosis	
	FCB	*FCBD*	*FCB*	*FCBD*
Sensitivity	98.48	98.94	96.97	96.36
Specificity	97.67	98.33	94.87	93.59
Accuracy	98.10 ± 0.6	98.65 ± 1.2	95.83 ± 1.2	94.86 ± 1.6
AUC	0.9810	0.9915	0.9745	0.9657

Table 6. Classification accuracy for keratosis vs. nodules and paralysis vs. non-paralysis

Performance Measures	Keratosis and Nodules		Paralysis and Others	
	FCB	*FCBD*	*FCB*	*FCBD*
Sensitivity	97.44	96.54	98.59	99.20
Specificity	96.67	95.50	98.12	99.06
Accuracy	97.10±1.2	96.09±1.4	98.36 ± 0.7	99.13 ± 0.8
AUC	0.9842	0.9810	0.9985	1

The classification results of adductor with nodules and keratosis are presented in Table 5. The accuracy of adductor vs. nodules with the FCB is 98.10% and the FCBD is 98.65%. Furthermore, the accuracy of adductor vs.

keratosis with the FCB is 95.83% and the FCBD is 94.86%. The results of keratosis with nodules and paralysis with all the other disorders are listed in Table 6. The best accuracy for keratosis vs. nodules is 97.10%, and paralysis vs. non-paralysis is 99.13%.

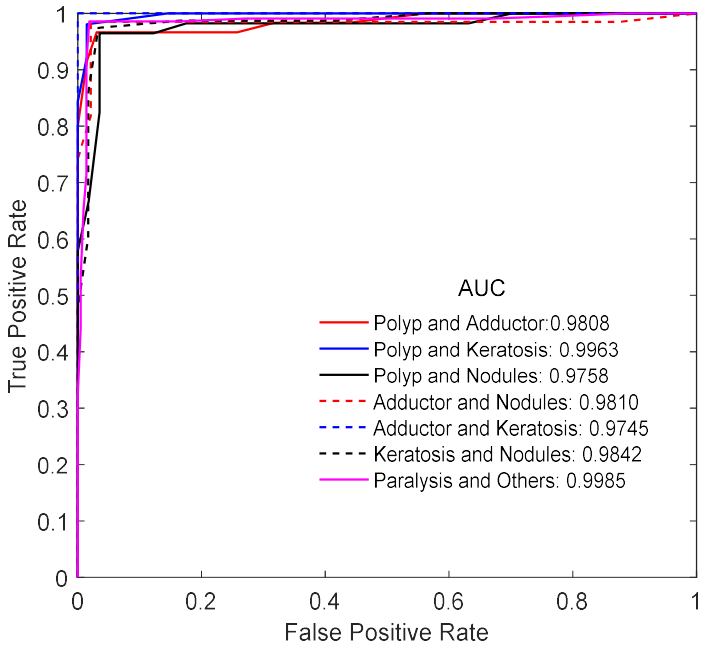


Figure 7. ROC curves and AUC for all experiments of disorder classification (Ex1 to Ex7).

It is very important for an automatic system to be reliable in the decision. To observe the reliability of the system for disorder classification, the AUC for all experiments is computed. For all classification experiments, the AUC is greater than 0.97, as shown in Fig. 7. The AUC is close to 1 for all experiments which suggest that the proposed system is reliable in the classification of various types of disorders. Therefore, it can be inferred that the proposed system can be deployed in smart homes and cities used reliably for classification of disorder remotely.

IV. DISCUSSION

Voice pathologies appear on the surface of the vocal folds because of the excessive use of the voice, smoking, drinking alcohol, and dehydration. Due to these pathologies, vocal folds experience abnormal behavior in vibration and cannot be opened and closed at regular intervals. This behavior of the vocal folds produces weaker, breathier, harsh, and strained voices. The signals of such voices contain noisy components due to the disorder of the vocal folds. Therefore, these voices feel unpleasant to the human ear. This is the reason why the human hearing system is simulated in this study to differentiate between different types of disorders and normal subjects.

To examine the speech signal produced by a normal and a disordered subject, an acoustic analysis is performed. The major component of the acoustic analysis is critical bandwidth phenomena, which simulate the human hearing system. The output of the acoustic analysis is the FCB and it is computed for each frame of the sample and 29 features are extracted for each frame. The interpretation of multidimensional FCB is not possible by the human mind. Therefore, a pattern recognition phase by using GMM is introduced for automatic assessment of disorders and its LL

values are used for the decision. The FCB with GMM provide quantitative analysis of the proposed system, and experimental results show that the system is accurate and reliable in the assessment of disorders.

A numerical value supported by a visual indication can enhance the accuracy of the diagnosis. Therefore, an intelligent system must be evaluated through qualitative analysis to observe its capability of decision without a classifier. To investigate the proposed system qualitatively, the computed FCB which can also be referred as auditory processed spectrum is plotted for disordered and normal subjects and shown in Fig. 8(a) and 8(b), respectively. It can be observed from Fig. 8(a) that the spectrum of the disordered subject experiences very frequent voice breaks due to the malfunctioning of the vocal folds. A significant difference between the spectrums of the disordered and normal subjects can be noted in the first 10 filters. It concludes that the visual clue provided by the acoustic analysis of the proposed system can be used for the assessment of disorders.

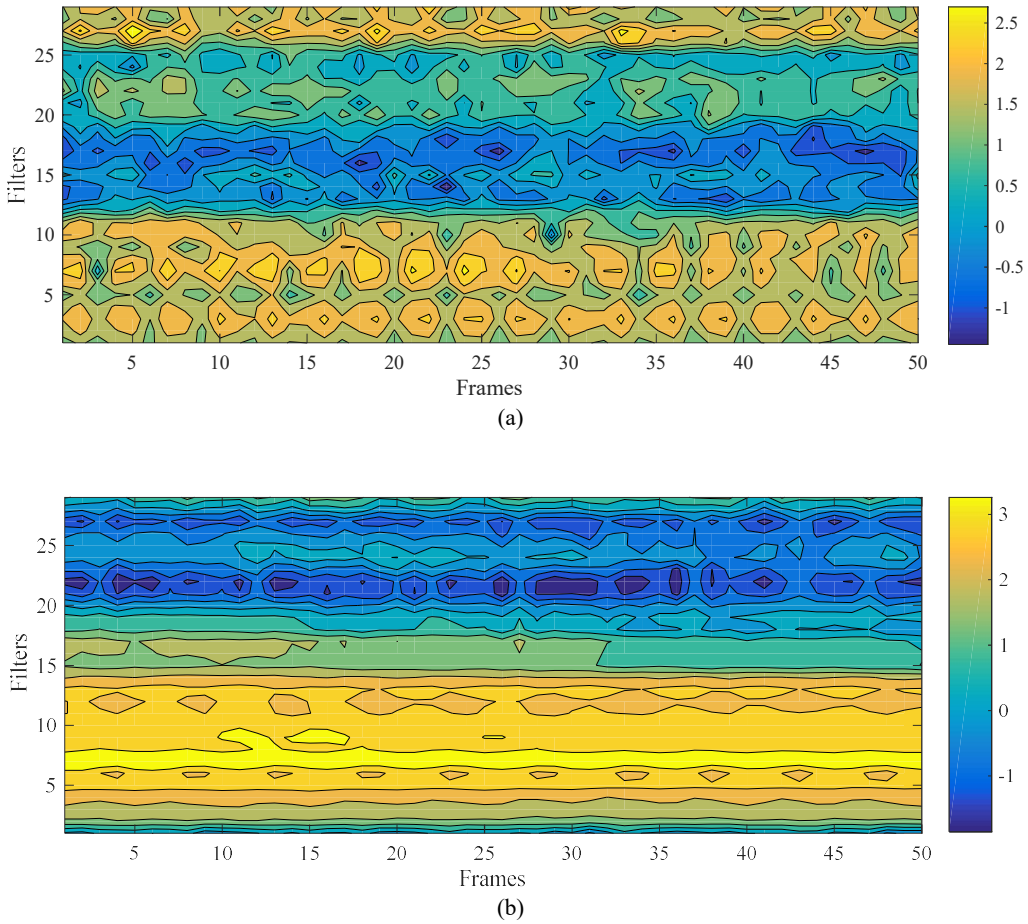


Figure 8. Energy contours in FCB for (a) pathological sample (b) normal sample.

The acoustic analysis of the voice can be accomplished through different kinds of speech features. In some acoustic analyses, long-term features such as shimmer and jitter are used for the assessment of pathological data [50-52], while other analyses have used short-term features [33, 53]. Most of the long-term features depend on a precise valuation of the fundamental frequency, which is itself a challenging task [41], particularly in pathological data. Therefore, short-term features are preferable over long-term features to obtain good accuracy. In the proposed system, the FCB is determined by short-term acoustic analysis.

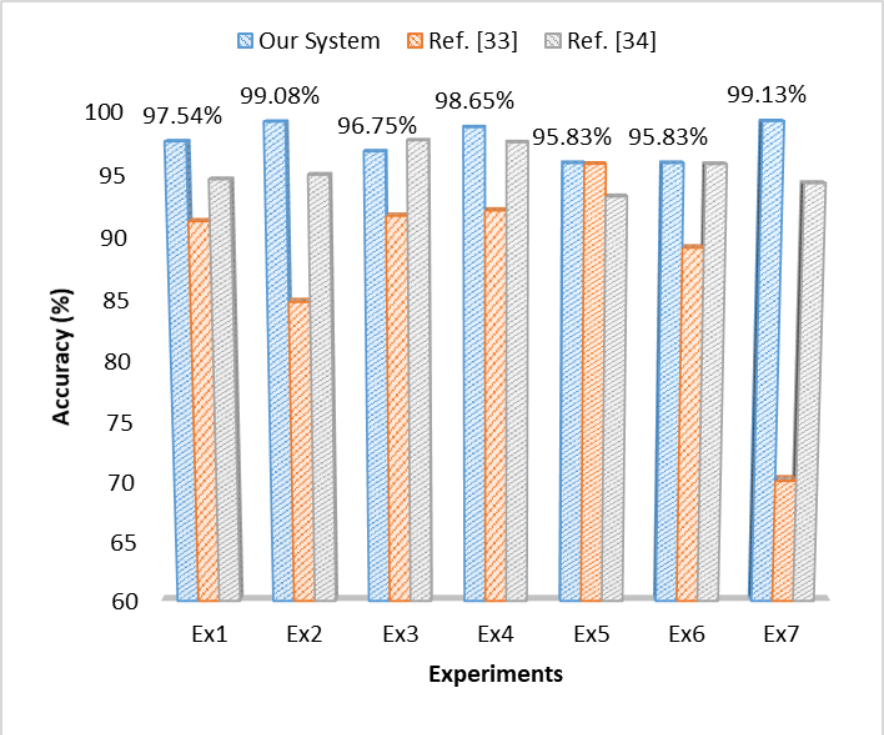


Figure 9. Comparison of the proposed system with the existing systems.

The setup for the classification of disorders is taken from [33, 34] to enable a comparison between our proposed system and the system presented in [33, 34]. The comparison is shown in Fig. 9. The accuracies of the systems [33, 34] are obtained from the respective studies, while, the accuracies of our system are taken from Tables 4 to 6. The proposed system obtained an accuracy of 97.54%, 99.08%, 96.75%, 98.65%, 95.83%, 95.83%, and 99.13% for the experiments Ex1, Ex2, Ex3, Ex4, Ex5, Ex6, and Ex7, respectively. The performance of our proposed classification system is better than the existing system.

The proposed intelligent system can contribute to healthcare industry significantly and has many potential applications. A large population around the world suffering from the vocal folds disorders which can be benefited from the proposed system. The general practitioner in the remote areas can use the system for early screening of the patients and can refer to a specialist for further consultation if the diagnosis of the system is positive. In addition, the proposed system can process the voice samples collected and transmitted by the IoT and retransmit the results of diagnosis. In this way, a patient can avoid the long waiting time and unwanted frequent visits to the hospitals and health centers.

V. CONCLUSION

An intelligent healthcare system for detection and classification of the vocal fold disorders is proposed and implemented in this study. The system can be installed in smart homes and cities for remote evaluation of voice samples. It can do the early screening of disorders to avoid the complications that may occur due to negligence or delay in diagnosis. The system computes the FCB features by using the critical bandwidth phenomena and gives to GMM based classifies for automatic decision. In addition, the FCB features provide the spectrums for the normal and disordered speech samples. The spectrum of a disordered subject highlights the voice breaks that occur due to vocal fold disorders. Normal and disordered subjects have significantly different patterns in the computed spectrums,

especially in the lower filters. The best obtained accuracy for the detection of disorders is 99.72%. The proposed system not only provided the best accuracy but also presented a clear visual indication for the presence of voice disorders. Furthermore, a classification accuracy of 97.54%, 99.08%, 96.75%, 98.65%, 95.83%, 95.83%, and 99.13% is achieved to differentiate between vocal fold polyp, keratosis, vocal fold paralysis, vocal fold nodules, and adductor spasmodic dysphonia. The performance of the proposed system is compared with existing systems, and it is found better.

The classification of disorders in the proposed system is binary in nature. The system determines the type of a disorder from two given disorders. However, the system can be modified to determine a certain type of disorder when all disorders are given.

ACKNOWLEDGMENTS

CONFLICTS OF INTEREST

The authors do not have any conflicts of interest.

REFERENCES

[1] V. McKelvey, "Spending more on in-home care," vol. Retrievd on March 1, 2017 from http://www.aarp.org/relationships/caregiving/info-01-2010/spending-more-on-in-home-care.html, 2010.

[2] United Nations. (2015). *World Population Ageing. Available at http://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015_Report.pdf*.

[3] Y. Mehmood, F. Ahmad, I. Yaqoob, A. Adnane, M. Imran, and S. Guizani, "Internet-of-Things-Based Smart Cities: Recent Advances and Challenges," *IEEE Communications Magazine,* vol. 55, pp. 16-24, 2017.

[4] M. S. Hossain, M. A. Rahman, and G. Muhammad, "Cyber–physical cloud-oriented multi-sensory smart home framework for elderly people: An energy efficiency perspective," *Journal of Parallel and Distributed Computing,* 2016.

[5] U. Aguilera, O. Peña, O. Belmonte, and D. López-de-Ipiña, "Citizen-centric data services for smarter cities," *Future Generation Computer Systems,* 2016.

[6] Z. Ali, G. Muhammad, and M. F. Alhamid, "An Automatic Health Monitoring System for Patients Suffering From Voice Complications in Smart Cities," *IEEE Access,* vol. 5, pp. 3900-3908, 2017.

[7] M. S. Hossain, "Patient status monitoring for smart home healthcare," in *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2016, pp. 1-6.

[8] G. Muhammad, "Automatic speech recognition using interlaced derivative pattern for cloud based healthcare system," *Cluster Computing,* vol. 18, pp. 795-802, 2015.

[9] G. Muhammad, T. A. Mesallam, K. H. Malki, M. Farahat, M. Alsulaiman, and M. Bukhari, "Formant analysis in dysphonic patients and automatic Arabic digit speech recognition," *Biomed Eng Online,* vol. 10, pp. 1-12, 2011.

[10] Z. Ali, I. Elamvazuthi, M. Alsulaiman, and G. Muhammad, "Detection of Voice Pathology using Fractal Dimension in a Multiresolution Analysis of Normal and Disordered Speech Signals," *Journal of Medical Systems,* vol. 40, pp. 1-10, 2015.

[11] N. Bhattacharyya, "The prevalence of voice problems among adults in the United States," *The Laryngoscope,* vol. 124, pp. 2359-2362, 2014.

[12] K. H. Malki, S. F. Al-Habib, A. A. Hagr, and M. M. Farahat, "Acoustic analysis of normal Saudi adult voices," *Saudi Med J,* vol. 30, pp. 1081-6, Aug 2009.

[13] J. Jiang, H.-J. Chen, J. Stern, and N. P. Solomon, "Vocal Efficiency Measurements in Subjects with Vocal Polyps and Nodules: A Preliminary Report," *Annals of Otology, Rhinology & Laryngology,* vol. 113, pp. 277-282, 2004.

[14] R. H. Martins, J. Defaveri, M. A. Domingues, and R. de Albuquerque e Silva, "Vocal polyps: clinical, morphological, and immunohistochemical aspects," *J Voice,* vol. 25, pp. 98-106, 2011.

[15] A. I. R. Fontes, P. T. V. Souza, A. Neto, #xe3, o. D. D., A. d. M. Martins*, et al.*, "Classification System of Pathological Voices Using Correntropy," *Mathematical Problems in Engineering,* vol. 2014, p. 7, 2014.

[16] T. Mau, "Diagnostic evaluation and management of hoarseness," *Med Clin North Am,* vol. 94, pp. 945-60, 2010.

[17] J. T. Cohen, K. K. Bach, G. N. Postma, and J. A. Koufman, "Clinical manifestations of laryngopharyngeal reflux," *Ear Nose Throat J,* vol. 81, pp. 19-23, 2002.

[18] L. H. Rosenthal, M. S. Benninger, and R. H. Deeb, "Vocal fold immobility: a longitudinal analysis of etiology over 20 years," *Laryngoscope,* vol. 117, pp. 1864-70, 2007.

[19] R. Leonard, "Voice therapy and vocal nodules in adults," *Curr Opin Otolaryngol Head Neck Surg,* vol. 17, pp. 453-457, 2009.

[20] K. Simonyan, F. Tovar-Moll, J. Ostuni, M. Hallett, V. F. Kalasinsky, M. R. Lewin-Smith*, et al.*, "Focal white matter changes in spasmodic dysphonia: a combined diffusion tensor imaging and neuropathological study," *Brain,* vol. 131, pp. 447-59, 2008.

[21] M. P. Cannito, M. Doiuchi, T. Murry, and G. E. Woodson, "Perceptual Structure of Adductor Spasmodic Dysphonia and Its Acoustic Correlates," *Journal of Voice,* vol. 26, pp. 818.e5-818.e13, 2012.

[22] K. H. Malki and T. A. Mesallam, "Psychosocial assessment of voice problems among Saudi teachers," *J Otolaryngol Head Neck Surg,* vol. 41, pp. 189-99, Jun 1 2012.

[23] 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 Hear Res,* vol. 47, pp. 281-93, 2004.

[24] N. Gori, H. Kadakia, V. Kashid, M. P. Hatode, and S. Natarajan, "Detection and Analysis of Microanuerysm in Diabetic Retinopathy using Fundus Image Processing," 2017.

[25] G. Tradigo, B. Calabrese, M. Macri, E. Vocaturo, N. Lombardo, and P. Veltri, "Voice signal features analysis and classification: looking for new diseases related parameters," in *Computational Biology and Health Informatics, 6th ACM Conference on Bioinformatics*, 2015, pp. 589-596.

[26] H. Wang and W. Hu, "Optimization of Pathological Voice Feature Based on KPCA and SVM," in *Biometric Recognition*. vol. 8833, Z. Sun, S. Shan, H. Sang, J. Zhou, Y. Wang, and W. Yuan, Eds., ed: Springer International Publishing, 2014, pp. 394-403.

[27] Z. Ali, I. Elamvazuthi, M. Alsulaiman, and G. Muhammad, "Detection of Voice Pathology using Fractal Dimension in a Multiresolution Analysis of Normal and Disordered Speech Signals," *Journal of Medical Systems,* vol. 40, p. 20, November 03 2015.

[28] Z. Ali, M. Talha, and M. Alsulaiman, "A Practical Approach: Design and Implementation of a Healthcare Software for Screening of Dysphonic Patients," *IEEE Access,* vol. 5, pp. 5844-5857, 2017.

[29] I. R. Titze, *Workshop on acoustic voice analysis: Summary statement*: National Center for Voice and Speech, 1995.

[30] M. P. Karnell, K. D. Hall, and K. L. Landahl, "Comparison of fundamental frequency and perturbation measurements among three analysis systems," *Journal of Voice,* vol. 9, pp. 383-393, 1995.

[31] Y. Zhang, J. J. Jiang, S. M. Wallace, and L. Zhou, "Comparison of nonlinear dynamic methods and perturbation methods for voice analysis," *J Acoust Soc Am,* vol. 118, pp. 2551-60, 2005.

[32] H. Cordeiro, C. Meneses, and J. Fonseca, "Continuous Speech Classification Systems for Voice Pathologies Identification," in *Technological Innovation for Cloud-Based Engineering Systems*. vol. 450, L. M. Camarinha-Matos, T. A. Baldissera, G. Di Orio, and F. Marques, Eds., ed: Springer International Publishing, 2015, pp. 217-224.

[33] 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.

[34] M. Markaki and Y. Stylianou, "Voice Pathology Detection and Discrimination Based on Modulation Spectral Features," *Audio, Speech, and Language Processing, IEEE Transactions on,* vol. 19, pp. 1938-1948, 2011.

[35] P. L. Dhingra and S. Dhingra, *Diseases of ear, nose and throat*, 6 ed.: Elsevier, India, 2014.

[36] Massachusette Eye & Ear Infirmry Voice & Speech LAB, "Disordered Voice Database Model 4337 (Ver. 1.03) ", ed. Lincoln Park, NJ: Kay Elemetrics Corp., 1994.

[37] T. Villa-Canas, E. Belalcazar-Bolamos, S. Bedoya-Jaramillo, J. F. Garces, J. R. Orozco-Arroyave, J. D. Arias-Londono*, et al.*, "Automatic detection of laryngeal pathologies using cepstral analysis in Mel and Bark scales," in *XVII Symposium of Image, Signal Processing, and Artificial Vision (STSIVA), 2012* 2012, pp. 116-121.

[38] J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, and G. Castellanos-Domínguez, "An improved method for voice pathology detection by means of a HMM-based feature space transformation," *Pattern Recognition,* vol. 43, pp. 3100-3112, 2010.

[39] M. Markaki and Y. Stylianou, "Voice Pathology Detection and Discrimination Based on Modulation Spectral Features," *IEEE Transactions on Audio, Speech, and Language Processing,,* vol. 19, pp. 1938-1948, 2011.

[40] 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 Transactions on Biomedical Engineering,,* vol. 53, pp. 1943-1953, 2006.

[41] V. Parsa and D. G. Jamieson, "Identification of Pathological Voices Using Glottal Noise Measures," *Journal of Speech, Language, and Hearing Research,* vol. 43, pp. 469-485, 2000.

[42] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proceedings of the IEEE,* vol. 66, pp. 51-83, 1978.

[43] E. Zwicker, "Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen)," *The Journal of the Acoustical Society of America,* vol. 33, pp. 248-248, 1961.

[44] C. M. Bishop, *Pattern Recognition and Machine Learning*: Springer-Verlag New York, 2006.

[45] J. Yang, X. Yuan, X. Liao, P. Llull, D. J. Brady, G. Sapiro*, et al.*, "Video Compressive Sensing Using Gaussian Mixture Models," *Image Processing, IEEE Transactions on,* vol. 23, pp. 4863-4878, 2014.

[46] Z. Ali, M. Alsulaiman, G. Muhammad, I. Elamvazuthi, and T. A. Mesallam, "Vocal fold disorder detection based on continuous speech by using MFCC and GMM," in *GCC Conference and Exhibition (GCC), 7th IEEE*, 2013, pp. 292-297.

[47] T. H. Falk and C. Wai-Yip, "Nonintrusive speech quality estimation using Gaussian mixture models," *Signal Processing Letters, IEEE,* vol. 13, pp. 108-111, 2006.

[48] S. Z. Selim and M. A. Ismail, "K-Means-Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. PAMI-6, pp. 81-87, 1984.

[49] R. A. Redner and H. F. Walker, "Mixture Densities, Maximum Likelihood and the EM Algorithm," *SIAM Review,* vol. 26, pp. 195-239, 1984.

[50] A. Al-nasheri, Z. Ali, G. Muhammad, M. Alsulaiman, K. H. Almalki, T. A. Mesallam*, et al.*, "Voice pathology detection with MDVP parameters using Arabic voice pathology database," in *Information Technology: Towards New Smart World (NSITNSW), 2015 5th National Symposium on*, 2015, pp. 1-5.

[51] 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," *Journal of Voice,* vol. 25, pp. e275-e289, 2011.

[52] K. Werth, D. Voigt, M. Dollinger, U. Eysholdt, and J. Lohscheller, "Clinical value of acoustic voice measures: a retrospective study," *Eur Arch Otorhinolaryngol,* vol. 267, pp. 1261-71, Aug 2010.

[53] L. F. Brinca, A. P. F. Batista, A. I. Tavares, I. C. Gonçalves, and M. L. Moreno, "Use of Cepstral Analyses for Differentiating Normal From Dysphonic Voices: A Comparative Study of Connected Speech Versus Sustained Vowel in European Portuguese Female Speakers," *Journal of Voice,* vol. 28, pp. 282-286, 5// 2014.