

# An IoT-based Smart Healthcare System to Detect Dysphonia

Zulfiqar Ali

School of Computer Science and Electronic Engineering, University of Essex, United Kingdom.

E-mail: [z.ali@essex.ac.uk](mailto:z.ali@essex.ac.uk)

Muhammad Imran

College of Applied Computer Science, King Saud University, Saudi Arabia.

E-mail: [cimran@ksu.edu.sa](mailto:cimran@ksu.edu.sa)

Mohammad Shoaib

College of Computer and Information Sciences, King Saud University, Saudi Arabia.

E-mail: [muhshoaib@ksu.edu.sa](mailto:muhshoaib@ksu.edu.sa)

## Abstract

Smart healthcare systems for the Internet of Things (IoT) platform are cost-efficient and facilitate continuous remote monitoring of patients to avoid unnecessary hospital visits and long waiting times to see practitioners. Presenting a smart healthcare system for the detection of dysphonia can reduce the suffering and pain of patients by providing an initial evaluation of voice. This preliminary feedback of voice could minimize the burden on ENT specialists by referring only genuine cases to them as well as giving an early alarm of potential voice complications to patients. Any possible delay in the treatment and/or inaccurate diagnosis using the subjective nature of tools may lead to severe circumstances for an individual because some types of dysphonia are life-threatening. Therefore, an accurate and reliable smart healthcare system for IoT platform to detect dysphonia is proposed and implemented in this study. Higher-order directional derivatives are used to analyze the time-frequency spectrum of signals in the proposed system. The computed derivatives provide essential and vital information by analyzing the spectrum along different directions to capture the changes that appeared due to malfunctioning the vocal folds. The proposed system provides 99.1% accuracy, while the sensitivity and specificity are 99.4% and 98.1%, respectively. The experimental results showed that the proposed system could provide better classification accuracy than the traditional non-directional first-order derivatives. Hence, the system can be used as a reliable tool for detecting dysphonia and implemented in edge devices to avoid latency issues and protect privacy, unlike cloud processing.

**Keywords** Edge analytics. higher-order local derivatives. vocal folds disorders. spectrum analysis. support vector machine (SVM).



## 1. Introduction

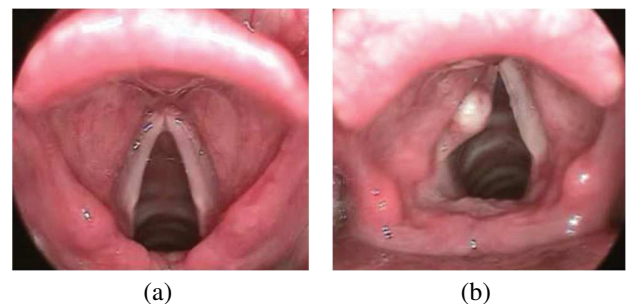
Recent developments in wireless communications and networking provide a solid platform for building smart homes and cities [1], where Smart health facilities will be one of the primary services for the quality and healthy life of residents [2,3]. Several smart health systems have been proposed to provide a higher standard of care by online monitoring of various diseases [4-7]. Different approaches and their benefits are summarized in a recent survey on online monitoring, detection, and support of the diagnosis of cardiovascular diseases [8]. In [9], a reliable smart healthcare system for the supervision of cardiomyopathy patients is presented to capture early symptoms of this disease. A recent study and a survey on brain tumor classification have highlighted the need for smart healthcare systems for remote monitoring of patients [10-12]. The Internet of Things (IoT) based smart healthcare systems efficiently collect and process patients' data [13,14] and automatically generate a diagnostic report. Such systems keep the patients and practitioners up-to-date about the health conditions to reduce the severe consequences of dysphonia delay/negligence [15,16].

According to the medical dictionary [17], dysphonia refers to a difficulty in speaking, usually evidenced by hoarseness. Hoarseness represents any deviation of voice quality as perceived by self or others [18]. The normal vocal folds and the ones affected by dysphonia (also refer as vocal folds disorders) are shown in Fig. 1. A large population, approximately one-third of the world population, is suffering from voice-related complications [19,20]. According to National Institute on Deafness and Other Communication Disorders [21], approximately 17.5 million people around U.S. are affected by voice problems. Any person can be affected by a vocal folds disorder; however, professional users of voice such as teachers, lawyers, and singers have high risks of vocal folds disorders [20,22]. The screening of vocal folds disorders is a crucial step in clinical treatment [23]. The clinical appraisal, also known as a subjective assessment, can be carried out using an endoscopic examination of the vocal folds and various acoustic and perceptual measurements [24]. The clinicians frequently use these techniques in their practice, but there are some limitations due to the subjective assessment nature. The limitations may be practitioner's experience and area of expertise, dysphonia rating scale, and vocal folds disorder degree [25]. Besides, some visualization tools, such as video laryngostroboscopy (LSC), are also used to detect vocal folds disorders by inspecting the vibration of the vocal folds [26]. The subjective interpretation of the LSC examination results strongly depends on the area of

specialization and professional expertise of an examiner, which presents the limitations of LSC examination. Therefore, some rating scales are introduced in literature [27,28]. However, there is no available standard approach for LSC examination due to the dependency on existing methods on pitch estimation, periodicity, sustained phonation frequency etc. [29,30].

Moreover, videoendoscopic high-speed imaging and videokymography have shown more accurate performance than video LSC [31], and they gained much interest recently [32-35]. Due to expensive equipment and intensive labor to analyze the data, practitioner and investigators are hesitant to adopt these new methods [36]. These limitations associated with the exiting subjective methods can be overcome through the automatic screening of vocal folds disorders.

For automatic screening, a smart healthcare system capable of collecting the voice of an individual using IoT and process it with a reliable and accurate system is essential. Various acoustic measures, including shimmer, jitter, signal-to-noise ratio, noise measures, and several frequencies and amplitude perturbation measures, have been used to screen vocal folds disorder [38] automatically.



**Fig. 1** (a) Normal vocal folds (b) Vocal folds affected by a voice pathology called a cyst. Various vocal folds pathologies are defined in [37]

The Multi-Dimensional Voice Program (MDVP) [38] is the most popular commercial software to perform the acoustic analysis of a speech signal to detect the disorders. Other related programs, such as CSpeech [39] and PRAAT [40], are also available to compute perturbation and noise measures for vocal folds disorder screening. The MDVP could calculate 33 different parameters for acoustic analysis, and the list of these parameters is provided in [41]. Arjmandi et al. [41] considered 22 MDVP parameters to conduct research for the differentiation of normal and disordered signals. The experiments were based on 100 subjects of Massachusetts Eye & Ear Infirmary (MEEI) vocal folds disorder database

[42] and contained 50 dysphonic patients and 50 normal subjects. The unselected parameters either do not reflect the voice quality or they are missing in the MEEI database. The same company distributes MDVP and MEEI database (Kay Elemetrics Corporation, Lincoln Park, NJ), and for some files in the MEEI database, all 33 parameters are not provided. As a result, authors used those 22 parameters that are relevant to voice quality and available for all files in the MEEI database. The parameters were given to six different classifiers to discriminate between normal and disordered subjects (i.e., quadratic discriminant classifier, nearest mean classifier, Parzen classifier, K-means nearest neighbor classifier, support vector machine (SVM), and multilayer perceptron (MLP)). The highest accuracy achieved was 89.29 for SVM, while the lowest accuracy was 78.90% and obtained with the quadratic discriminant classifier. Moreover, two feature reduction techniques and six feature selection methods were implemented to improve the accuracy of the method. The method became complicated when using many algorithms and classifiers, however, an enhanced accuracy could not be achieved. The accuracy was improved by 1.71% with feature selection methods and 4.97% with feature reduction methods.

Among these long-term parameters, perturbation measures, jitter, and shimmer are used frequently. However, some studies have suggested that these measures are not consistent, and unreliable results could be obtained as a result Pepered et al. [43] reported that the measurement of jitter for nodules and normal subjects was significantly different; however, shimmer and normal subjects did not show significant differences when compared to nodule subjects. In a similar way, Lin et al. [44] concluded that the performance of percent jitter to differentiate between the group of vocal folds disorder subjects and normal subjects was good, while the ability of shimmer to discriminate between normal and disordered subjects was discouraging. On the contrary, Rosen et al. [45] found that the significance of shimmer to detect the nodules and normal subjects is higher than those of jitter. In a recent study [46], it was reported that when a speech signal is non-periodic, then it is not possible to extract the MDVP parameters accurately because shimmer and jitter are highly dependents on the fundamental frequency, which cannot be calculated precisely for non-periodic signals.

It can be observed from the existing studies that perturbation and noise measures could not provide good results for the detection of vocal folds disorders. Therefore, nonlinear dynamic features are also used with perturbation and noise measures to enhance the performance of the vocal folds disorder detection methods. In [47], fractal dimensions are combined with the MDVP parameters to improve the accuracy of the disorder detection system. At the same time, a method based on Local Binary Pattern (LBP) operator is proposed to overcome the problem of non-periodicity [48]. One of the issues regarding this method is that it only works for non-periodic signals. Besides, the LBP codes [49] do not provide detailed information in a specific direction. Therefore, higher-order local derivatives are used in this study to develop a high-performance smart healthcare system to detect dysphonia.

To the best of our knowledge, the higher-order local derivatives have never been used for the classification of normal and disordered speech signals. In this study, the higher-order derivatives analyze the time-frequency spectrum to determine the patterns along with various directions. Besides, The computed patterns provide distinctive and detailed information, unlike the traditional LBP codes. This information plays an important role in the classification of two types of signals. The experiments suggest that the information along different directions provides complementary information, and hence, the fusion at feature and decision level improved the classification results. In addition, the proposed system can detect vocal folds disorders in the presence of periodic as well as non-periodic signals.

The remaining part of the paper is organized as follows: Section II describes the proposed system and elaborated steps to determine the time-frequency spectrum and higher-order local derivatives. Section III presents the detailed experimental setup and results for the classification of normal and disordered speech signals using the first and second-order derivatives. In addition, this section also presents the results obtained by applying feature level and decision level fusion. In section IV, a comparison of the proposed system with the existing approaches is reported. Finally, Section V draws some conclusions and suggests future recommendations.

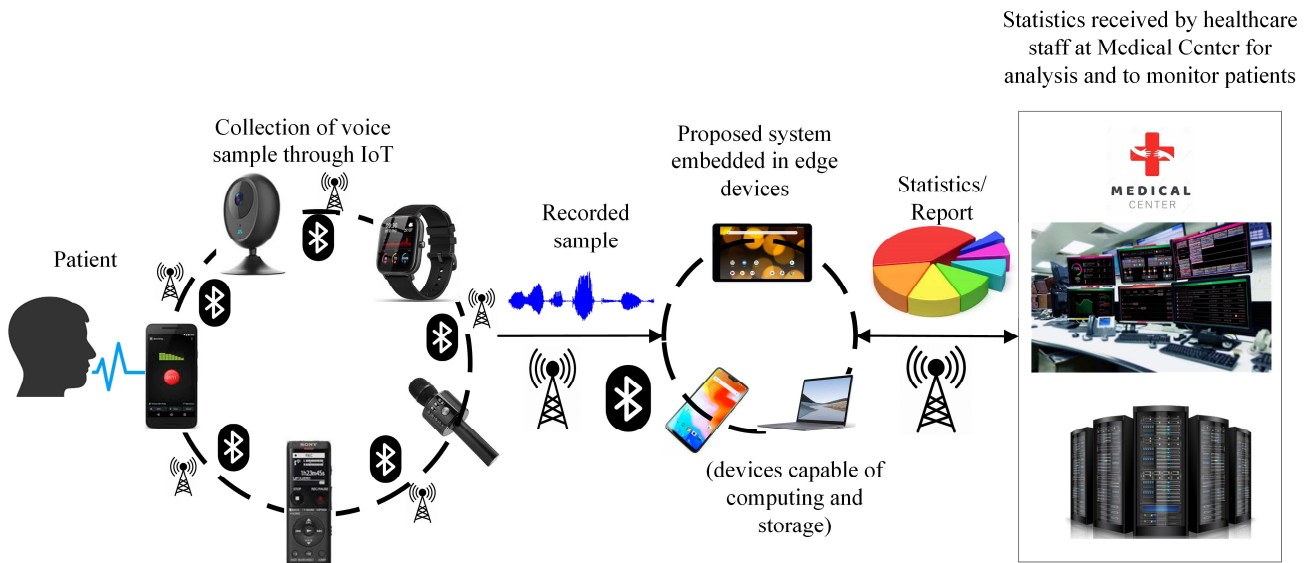


Fig. 2 IoT framework for the proposed healthcare system to detect vocal folds disorders

## 2. Proposed smart healthcare system for IoT platform

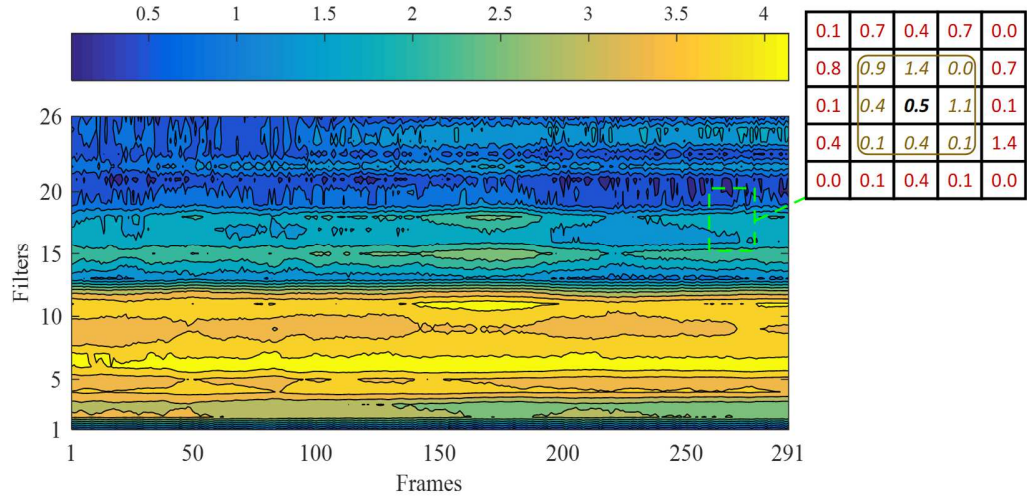
One of the main challenges of IoT based healthcare system is the central processing of the collected data using cloud computing. It not only raises the privacy and security issues of patients' data but also elevates the problem of latency [50]. These problems can be avoided by processing the data near the point of collection via moving the intelligence to local devices (edge devices) so that the data are not transmitted for central processing, and ultimately, there will be no issues of latency. These local devices can be laptops, smartphones, smartwatches, or tablets. The proposed smart healthcare system with IoT framework to detect dysphonia is presented in Fig. 2.

As shown in Fig. 2, various devices such as smartphones, voice recorders, webcams, and microphones can be used to record a voice sample of a patient. Then, the recorded sample is transmitted to edge devices through Bluetooth or Wi-Fi for the detection of vocal folds disorders using the proposed healthcare system. The proposed system is embedded in the edge devices (capable of computing and storage). The system evaluates the voice sample and sends the generated report to healthcare staff at a medical center through Wi-Fi. In this way, a health practitioner can monitor the condition of patients remotely, which is very crucial, especially after surgery, to eliminate any risk of complications.

To detect vocal folds disorders, the proposed system

investigates recorded speech signals. The vocal folds disorders exhibit irregular patterns in speech signals. Consequently, the signals become more complex and transient as compared to those of normal persons. Time-frequency analysis is performed using Mel-spaced bandpass filters to determine the trends in normal and disordered signals. This analysis provides a time-frequency spectrum that is further explored using higher-order local derivatives. The analysis of spectrum with higher-order derivatives highlights the changes in different directions and provides vital information to distinguish between normal and disordered signals. The computed information is passed to the SVM for automatic classification of both types of signals. The time-frequency spectrum of a speech signal is computed using the following steps:

- 1) A speech signal is divided into overlapping frames. The current frame starts from the middle of the previous frame to avoid the loss of any information at the ends.
- 2) Each frame is multiplied by a Hamming window to taper the frame ends close to zero. It causes continuity between successive frames and avoids spectral leakage during Fourier transformation (FT), which is an important component of time-frequency analysis.
- 3) Subsequently, FT is implemented on each overlapped frame to determine the contribution of each frequency component using the fast Fourier algorithm due to its low computational cost.

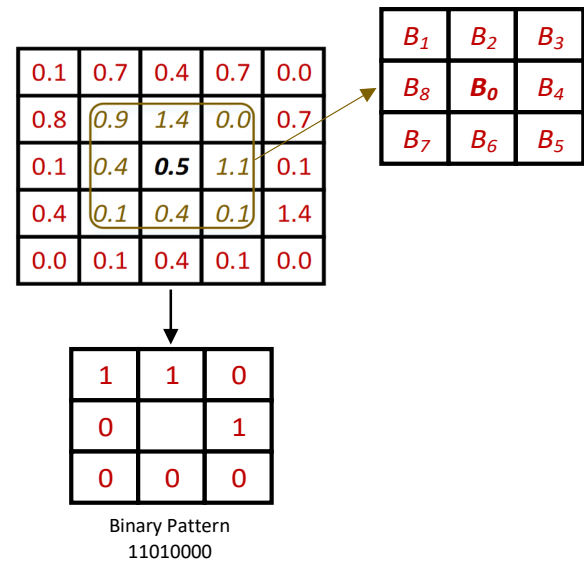


**Fig. 3** Time-frequency spectrum with a highlighted area to compute first and second-order derivatives

- 4) The spectrum obtained after applying FT is further processed to simulate the human auditory perception by passing the spectrum through Mel-spaced bandpass filters. The bandwidth of these filters is linear for lower frequencies (up to 1000 Hz) and logarithmic for higher frequencies. As a result, a time-frequency spectrum is obtained, as shown in Fig. 3.
- 5) Finally, the obtained time-frequency spectrum is divided into small blocks to analyze its texture using the first and second-order local derivatives. A block of dimension  $5 \times 5$  representing a portion of the spectrum is shown in Fig. 3.

To compute the first-order derivative, eight neighboring elements (taken anti-clockwise) of the center element are considered. In Fig. 4, the bold element  $B_0$  is a center of  $5 \times 5$  block  $B$ , while italic elements ( $B_1, B_2, B_3, B_4, B_5, B_6, B_7,$  and  $B_8$ , where 0.9 is  $B_1$ ) are its neighbors. The first-order derivative descriptor is also referred to as a Local Binary Pattern (LBP), and it provides a non-directional derivative pattern. The computation of the first-order derivative is shown in Fig. 4. The center element is compared with all neighbors; when a neighbor is greater than the center element, then it is replaced by 1; otherwise, 0. The criterion is given in Eq. 1, where  $B'_i$  represents the first-order derivative of the corresponding element. The computed first-order derivative in Fig. 4 is 11010000, where the element at the top-right corner indicates the most significant bit. A decimal number equivalent to the computed binary number is the required LBP code. Similarly, each element of the spectrum is replaced by the LBP code, and a histogram is obtained. The number of

bins in computed histogram is 256, where a bin represents the frequency of an LBP code.



**Fig. 4** Computation of the first-order derivative in a  $5 \times 5$  block of the spectrum

$$B'_i = \begin{cases} 1 & \text{if } B_i - B_0 > 0 \\ 0 & \text{if } B_i - B_0 \leq 0 \end{cases} \quad (1)$$

where  $i = 1, 2, 3, K, 8$ .

As the first-order derivative cannot provide detailed information, therefore, the second-order derivative is also

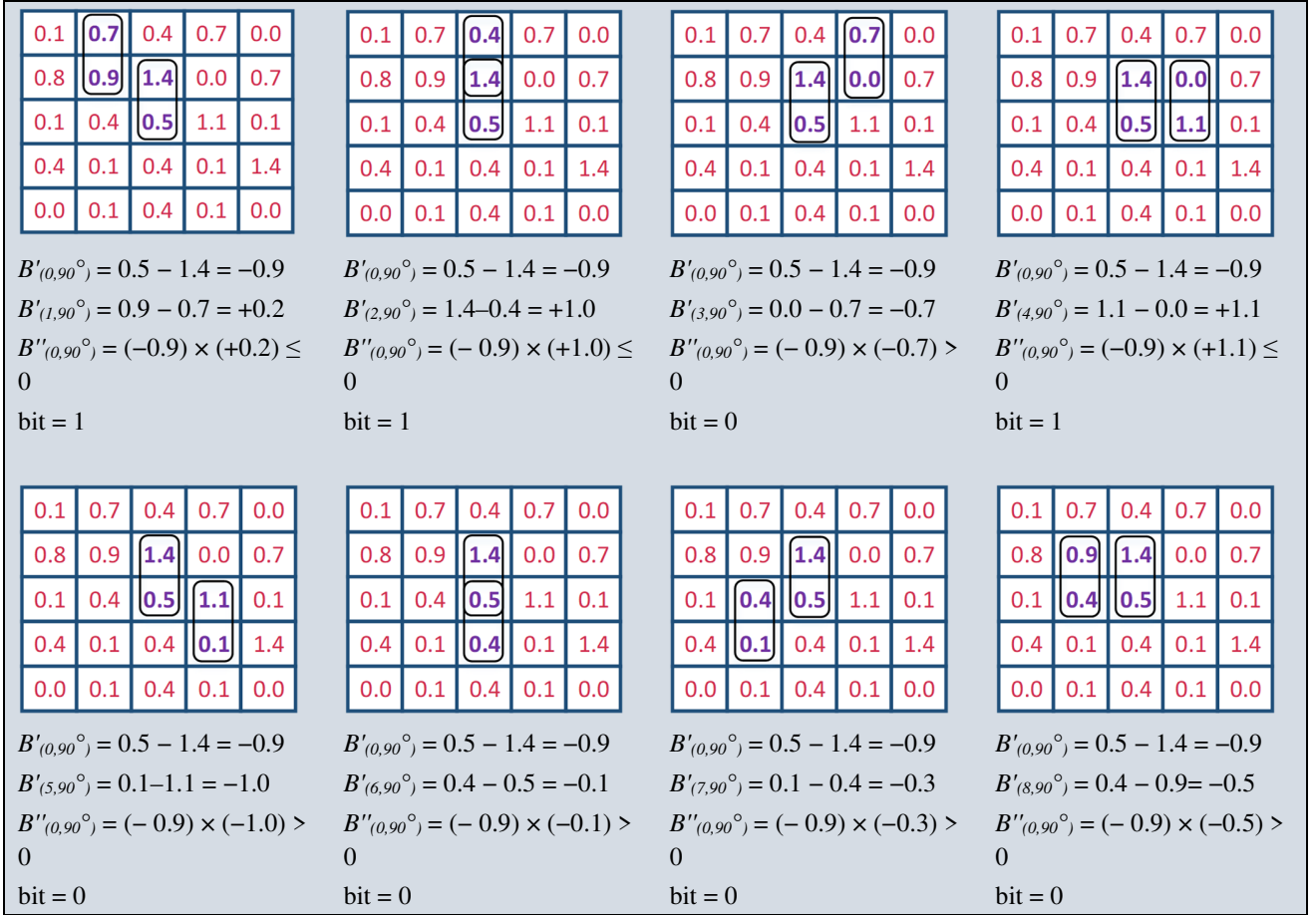


Fig. 5 Process of computing second-order derivative for  $B_0$  along  $90^\circ$

computed for each element of  $5 \times 5$  block  $B$ . It provides the change along with different directions, i.e.,  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$ . The calculation of the first-order derivative in all directions using the criteria mentioned in Eq. 2 was performed to compute the second-order derivative for a center element  $B_0$ .

$$\begin{aligned}
 B'_{(0,0^\circ)} &= B_0 - B_4 \\
 B'_{(0,45^\circ)} &= B_0 - B_3 \\
 B'_{(0,90^\circ)} &= B_0 - B_2 \\
 B'_{(0,135^\circ)} &= B_0 - B_1
 \end{aligned} \tag{2}$$

Then, computing the first derivative for all neighbors  $B_1$ ,  $B_2$ ,  $B_3$ ,  $B_4$ ,  $B_5$ ,  $B_6$ ,  $B_7$ , and  $B_8$  in all directions. The pattern of second-order derivative for  $B_0$  in a direction  $\gamma$  is given by Eq. 3.

The complete process to determine the second-order derivative for  $B_0$  along  $90^\circ$  is illustrated in Fig. 5. For instance,

to obtain the first bit of the second-order pattern, the first-order derivative for  $B_0$  and  $B_1$  are computed by subtracting the respective neighbors along  $90^\circ$  from them. Next, multiplying the derivatives; when the product is zero or negative, the required bit is 1; otherwise 0. Similarly, the bits for other neighbors  $B_2$ ,  $B_3$ ,  $B_4$ ,  $B_5$ ,  $B_6$ ,  $B_7$ , and  $B_8$  were also computed. The obtained 8-bits are 11010000. The second derivatives for  $B_0$  in the other three directions  $0^\circ$ ,  $45^\circ$ , and  $135^\circ$ , are 01010100, 00101111, and 11000110, respectively.

$$B''_{(0,\gamma)} = \begin{cases} 1 & \text{if } B'_{(0,\gamma)} \cdot B'_{(i,\gamma)} \leq 0 \\ 0 & \text{if } B'_{(0,\gamma)} \cdot B'_{(i,\gamma)} > 0 \end{cases}$$

where  $i = 1, 2, 3, \dots, 8$ . (3)

and  $\gamma = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ .

In this way, second derivative patterns are determined for each element of  $5 \times 5$  block in the time-frequency spectrum. As a result, a histogram along each direction could be



obtained. These histograms are transferred to SVM for automatic classification of normal and disordered signals.

### 3. Experimental setup and results

To validate the proposed smart healthcare system, many experiments are performed for the classification of normal and disordered speech signals, which are recorded using a microphone. The proposed system processes a speech signal to compute the time-frequency spectrum. The speech signal is divided into frames of length 512 samples with overlapping of 256 samples with the previous frames to compute the spectrum. The hamming window and FT are implemented with 512 points using 26 Mel-spaced bandpass filters to obtain the spectrum. During the texture analysis of the computed time-frequency spectrum, zero-padding is performed at the boundaries of the spectrum to avoid any information loss. In this way, the first and second-order derivatives are calculated for every element of the spectrum.

Four approaches are used to perform the classification of normal and disordered signals: Firstly, the first-order non-directional derivative (LBP codes) is transferred to SVM for automatic classification. Secondly, the second-order derivatives along each direction are given to SVM individually. Then, histograms of all directions are concatenated and given to SVM. This phenomenon is termed feature level fusion. Finally, the individual decisions of first- and second-order derivative (in all directions) are combined to reach the final consensus about the signal types. This approach is known as decision level fusion. The proposed system, with all approaches used for the experiments, is depicted in Fig. 6.

In each approach, the binary classifier SVM is implemented using three kernels linear, quadratic (quad), and radial basis function (RBF). The feature level fusion increases the overall dimension of the feature vector in which SVM performs relatively well compared to other classifiers based on clustering algorithms.

Various performance measures are used to examine the performance of the proposed system. These measures include sensitivity, specificity, and accuracy with the following definitions: sensitivity (SEN) provides the ratio of the truly detected disordered signals, specificity (SPE) gives the ratio of the correctly detected normal signal, and accuracy describes the truly detected disordered and normal signal in the complete dataset. SEN, SPE, and ACC are computed for each kernel of SVM in every experiment. All experiments are conducted using the speech signal of MEEI voice disorder database.

#### 3.1 MEEI voice disorder database

MEEI database [42] is a collection of normal and voice disordered speech signals. It was recorded at MEEI laboratory and publicly available through PENTAX Medical, USA. A subset of MEEI database containing 173 disordered and 53 normal signals is considered in this study to conduct all experiments. Among the 173 disordered subjects, 22 suffer from adductor, 20 from vocal nodules, 26 from keratosis, 20 from vocal fold polyp, and 85 from paralysis. Among the 53 normal subjects, 21 are males, and the remaining 32 are female. While 70 disordered subjects are males, and 103 are females. For pathological subjects, the ages of male patients are in the range of 26–58 years, while those for female patients are in the range of 21–51 years. The average age for men is 41.71 years, and for women, it is 37.58 years. Overall, the age range of patients is between 21 and 58 years, indicating that individuals of any age can be affected by vocal folds disorders [51].

This subset has been used in many studies [52,13,53]. The duration of each signal is three seconds and one second for normal and disordered signals, respectively. The possible reasons for the shorter duration of the disordered signals are the medical conditions of patients, including the severity of pain while speaking and the inability to hold breath for a long-time during the recording. All signals contain sustained vowels/ah/.

#### 3.2 Classification with first-order non-directional derivative

The non-directional first-order derivative compares the center element with all eight neighbors. Consequently, an eight-bit pattern of 0s and 1s is obtained, which describes the randomness in the time-frequency spectrum. The irregular vibration of the vocal folds due to voice disorders exhibits disruptions in the time-frequency spectrum. The first-order derivative captures such abnormalities in the spectrum. For instance, the spectrum of the disordered signal contains more 1-to-0 or 0-to-1 transitions compared to the spectrum of normal signals. The SVM differentiates the signals based on the computed pattern. The classification results using the first-order non-directional derivatives are provided in Table 1.

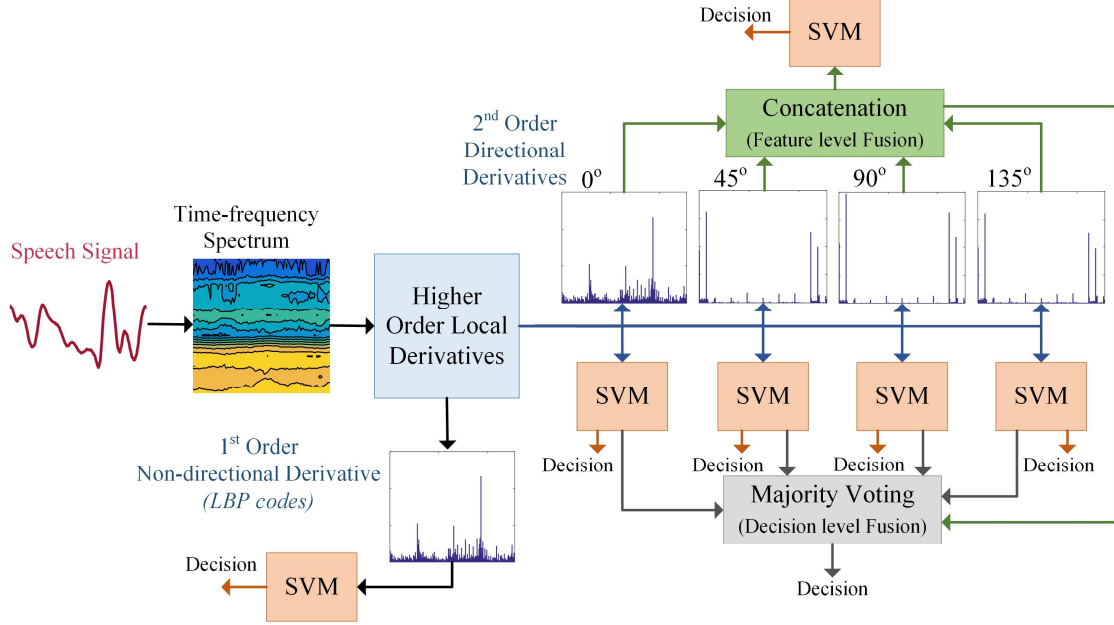
**Table 1** Classification Results using 1<sup>st</sup> order non-directional derivative (LBP codes)

Kernel	SEN	SPE	ACC
<i>Linear</i>	91.3	75.5	87.6
<i>Quad</i>	93.1	81.1	90.3



<i>RBF</i>	94.8	86.8	92.9
------------	------	------	------

The experimental results show that the patterns are not linearly separable. Therefore, the highest accuracy



**Fig. 6** Proposed smart healthcare system for classification of normal and disordered signal with all approaches

obtained was with RBF kernel, which is 92.9%. All measures are averaged over 5-folds. The standard deviation of the accuracy among 5-folds is 1.8.

### 3.3 Classification with second-order directional derivative

In contrast to the first-order non-directional derivatives determining the relationship between the neighbors and center element, the high order derivative captured the local information by highlighting the following distinctive information along with various directions.

- 1) The second-order derivative along  $0^\circ$  provides the change in frequency with time (frames).
- 2) The second-order derivative along  $90^\circ$  gives the variation in time (frames) with respect to frequency.
- 3) The second-order derivatives along  $45^\circ$  and  $135^\circ$  describe the variation when time and frequency change together. Along  $45^\circ$ , both frame and filter numbers increase, while along  $135^\circ$ , the filter number increases and the frame number decreases.

The classification results for second-order derivative along all directions are listed in Table 2. The results suggest that the

proposed system provides better classification accuracy when both time and frequency are changed simultaneously in comparison to the change in either time or frequency. The maximum obtained accuracy is 94.7%, with a standard deviation of 1.4 over 5-folds.

**Table 2** Classification results using 2<sup>nd</sup> order directional derivatives along  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$

Direction	Kernel	SEN	SPE	ACC
$0^\circ$	<i>Linear</i>	93.1	79.2	89.8
	<i>Quad</i>	94.2	84.9	92.0
	<i>RBF</i>	94.8	88.7	93.4
$45^\circ$	<i>Linear</i>	91.9	77.4	88.5
	<i>Quad</i>	93.6	83.0	91.2
	<i>RBF</i>	96.0	90.6	94.7
$90^\circ$	<i>Linear</i>	89.0	92.5	89.8
	<i>Quad</i>	92.5	94.3	92.9
	<i>RBF</i>	95.4	88.7	93.8
$135^\circ$	<i>Linear</i>	91.3	77.4	88.1
	<i>Quad</i>	94.2	83.0	91.6
	<i>RBF</i>	95.4	90.6	94.2

In Table 2, the variation in accuracy supports that the derivative along different directions may provide complementary information in the classification of signals. For example, a signal detected as a false positive in a certain direction might be detected as a true positive in another direction. The fusion of information can verify such situations.

### 3.4 Classification Results with Fusion

The information can be fused in two different ways: i) feature level fusion, and ii) decision level fusion. The feature level fusion is achieved by concatenating the histogram of all directions, while the decision level fusion is done by following the majority voting criteria on the individual decisions of all computed derivatives. The majority voting is based on the decision of each direction ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$ ) and those obtained by the fusion of these four directions, as shown in Fig. 6. In other words, the number of inputs for the majority voting is five.

The experimental results for both types of fusion are given in Table 3 and Table 4, respectively.

**Table 3** Classification results with feature level fusion (concatenation of directional derivatives of all directions)

Kernel	SEN	SPE	ACC
<i>Linear</i>	88.4	90.6	88.9

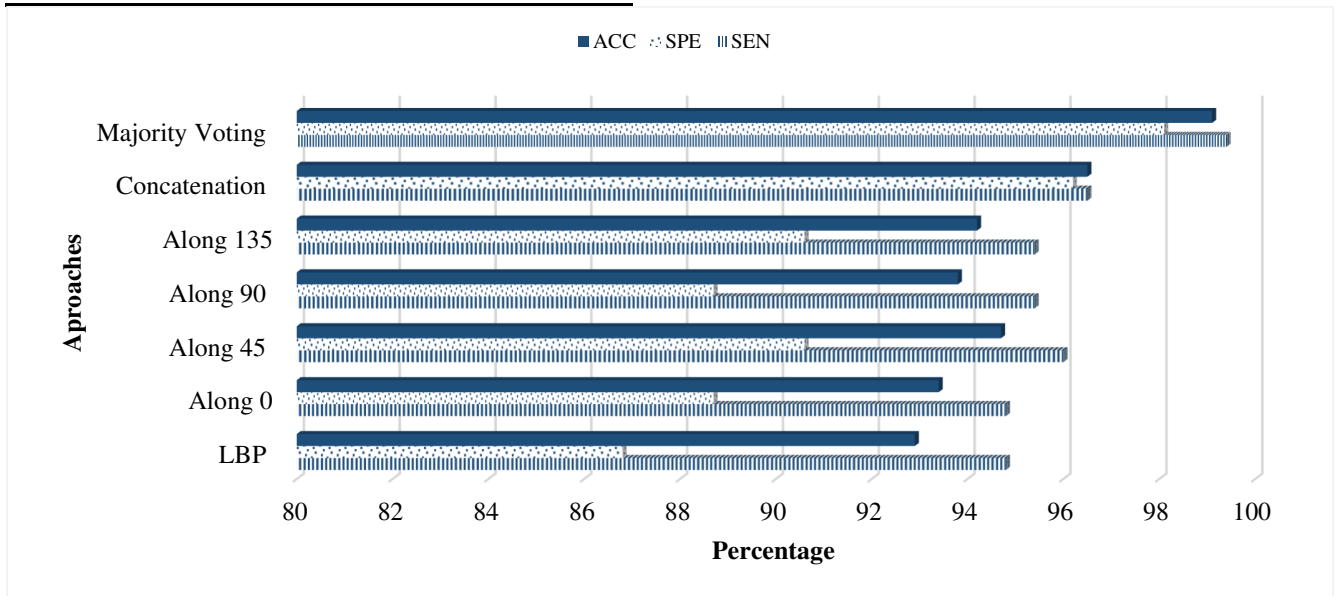
<i>Quad</i>	93.1	94.3	93.4
<i>RBF</i>	96.5	96.2	96.5

**Table 4** Classification results with decision level fusion using majority voting

Kernel	SEN	SPE	ACC
<i>Linear</i>	90.8	84.9	89.4
<i>Quad</i>	95.4	92.5	94.7
<i>RBF</i>	99.4	98.1	99.1

## 4. Summary and comparison

Four different approaches are used for the classification of normal and disordered speech signals to investigate the proposed healthcare system. In each approach, the highest accuracy is achieved with the RBF kernel. The reason is that the computed derivatives are not linearly separable, whereas the RBF kernel mapped the derivatives into high-dimensional space, which made the given data separable. The summary of the obtained results for all approaches is depicted in Fig. 7. The best obtained accuracy in a single direction is 94.7%, i.e., along  $45^\circ$ , when time and frequency change together in the time-frequency spectrum. In other approaches, feature level and decision level fusion achieved the highest accuracy of 96.5% and 99.1%, respectively. The improved accuracies in case of



**Fig. 7** Summary of the results of all approaches of the proposed healthcare system

fusion shows that derivatives along different directions provide complementary information for the detection of vocal folds disorders. For all approaches, 95% confidence interval is computed to observe the reliability of the decision of the proposed system. The obtained p-value is less than 0.05 in all approaches, which rejects the null hypothesis that SVM's decision values to distinguish the normal and disordered signal are from continuous distributions with equal medians. Therefore, the proposed system is reliable in decision making, i.e., the detection of dysphonia.

In a previous study [54], the LBP operator (non-directional first-order derivative) is used to compute LBP codes. Then, the computed codes are directly given to GMM for the classification of normal and disordered speech signals. The obtained accuracy was 92.50%, while SEN and SPE are 95% and 90%, respectively. In another study [48], the LBP descriptor is used for the recursion plot [55]. According to their work, a significant difference in the computed histograms of normal and disordered signals has been found because histograms of disordered signals contain a high number of non-uniform LBP codes than the normal signals. The non-uniform codes consist of more than two transitions from 0-to-1 or 1-to-0, indicating that the signal is transient and irregular. Only Type 2 and Type 3 signals are considered to perform the experiments in [48]. Consequently, the accuracy of the system is significantly improved. The system provided an accuracy of 97.73%. While, the achieved SEN and SPE are 98% and 96%, respectively. The accuracy is improved due to the exclusion of Type 1 disordered signals, which are closed to the normal signals and create confusion for the classifier.

Using the proposed system, an accuracy of 99.1% with SEN and SPE of 99.4% and 98.1%, respectively, have been obtained in this work. The results revealed that higher accuracy could be gained compared to existing systems [54]. Besides, an accuracy of 99.98% was achieved using MDVP parameters [46]. The obtained accuracy is high; however, MDVP parameters are not reliable for non-periodic signals [56] due to the irregular behavior of the vocal folds because voice pathology makes the signal complex and transient. As a result, non-periodicity appeared in the signal, especially when the severity of the disorder is high. Therefore, the approach used previously [46] cannot be used for non-periodic signals, and it is one of the major issues of the study. Moreover, other researchers [48] provide the solution for non-

periodic signals, but they cannot detect dysphonia accurately when the disordered signal is periodic.

However, the proposed system has the ability to detect dysphonia in the presence of all types of signals, i.e., periodic and non-periodic, which is a strong and significant aspect of the proposed system.

**Acknowledgments** The authors are thankful to the Deanship of Scientific Research, King Saud University Riyadh Saudi Arabia for funding through Research Group under Project no. RG-1439-036. The authors thank the Deanship of Scientific Research and RSSU at King Saud University for their technical support.

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## 5. Conclusions

An accurate and reliable healthcare system for IoT environment for the detection of vocal folds disorders is proposed and adopted in this study. The proposed system highlights the difference in the time-frequency spectrum of both normal and disordered signals using the higher-order derivatives, which encode the distinctive spatial information in a local region. The system is implemented using a machine-learning algorithm to make an automatic decision. Hence, it can be embedded into the edge devices in IoT environments to monitor patients remotely. By detecting vocal folds disorders at an early stage, patients' lives can be saved because some of the disorders are life-threatening in case of delay and/or negligence. Moreover, the patients can be observed remotely after the surgery of vocal folds to be followed up. Such edge-centric healthcare system avoids latency as well as the breach of patients' confidential health data as it processes the data near the collection point without transmission over a network. Therefore, no watermarking and cryptography solutions are required. Some suitable techniques, such as principal component analysis, will be implemented to determine the highly discriminant patterns in future work. Consequently, it will reduce the number of features, confirming the efficiency of the proposed system.

## References

1. Yang P, Stankevicius D, Marozas V, Deng Z, Liu E, Lukosevicius A, Dong F, Xu L, Min G (2018)

- Lifelogging data validation model for internet of things enabled personalized healthcare. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 48 (1):50-64. doi:10.1109/TSMC.2016.2586075
2. Guelzim T, Obaidat MS, Sadoun B (2016) Chapter 1 - Introduction and overview of key enabling technologies for smart cities and homes. In: *Smart Cities and Homes*. Morgan Kaufmann, Boston, pp 1-16. doi:<http://dx.doi.org/10.1016/B978-0-12-803454-5.00001-8>
  3. Raza M, Awais M, Singh N, Imran M, Hussain S (2020) Intelligent IoT Framework for Indoor Healthcare Monitoring of Parkinson's Disease Patient. *IEEE Journal on Selected Areas in Communications*:1-1. doi:10.1109/JSAC.2020.3021571
  4. Dourado CMJM, Silva SPPD, Nóbrega RVMD, Filho PPR, Muhammad K, Albuquerque VHCD (2020) An Open IoHT-based Deep Learning Framework for Online Medical Image Recognition. *IEEE Journal on Selected Areas in Communications*:1-1. doi:10.1109/JSAC.2020.3020598
  5. Naseer A, Rani M, Naz S, Razzak MI, Imran M, Xu G (2020) Refining Parkinson's neurological disorder identification through deep transfer learning. *Neural Computing and Applications* 32 (3):839-854. doi:10.1007/s00521-019-04069-0
  6. Ali F, El-Sappagh S, Islam SMR, Ali A, Attique M, Imran M, Kwak K-S (2021) An intelligent healthcare monitoring framework using wearable sensors and social networking data. *Future Generation Computer Systems* 114:23-43. doi:<https://doi.org/10.1016/j.future.2020.07.047>
  7. Ali F, El-Sappagh S, Islam SMR, Kwak D, Ali A, Imran M, Kwak K-S (2020) A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. *Information Fusion* 63:208-222. doi:<https://doi.org/10.1016/j.inffus.2020.06.008>
  8. Santos MAG, Munoz R, Olivares R, Filho PPR, Ser JD, Albuquerque VHCd (2020) Online heart monitoring systems on the internet of health things environments: A survey, a reference model and an outlook. *Information Fusion* 53:222-239. doi:<https://doi.org/10.1016/j.inffus.2019.06.004>
  9. Ding W, Abdel-Basset M, Eldrandaly KA, Abdel-Fatah L, Albuquerque VHCd (2020) Smart Supervision of Cardiomyopathy Based on Fuzzy Harris Hawks Optimizer and Wearable Sensing Data Optimization: A New Model. *IEEE Transactions on Cybernetics*:1-15. doi:10.1109/TCYB.2020.3000440
  10. Muhammad K, Khan S, Ser JD, Albuquerque VHCd (2020) Deep Learning for Multigrade Brain Tumor Classification in Smart Healthcare Systems: A Prospective Survey. *IEEE Transactions on Neural Networks and Learning Systems*:1-16. doi:10.1109/TNNLS.2020.2995800
  11. Rehman A, Naz S, Razzak MI, Akram F, Imran M (2020) A Deep Learning-Based Framework for Automatic Brain Tumors Classification Using Transfer Learning. *Circuits, Systems, and Signal Processing* 39 (2):757-775. doi:10.1007/s00034-019-01246-3
  12. Razzak MI, Imran M, Xu G (2019) Efficient Brain Tumor Segmentation With Multiscale Two-Pathway-Group Conventional Neural Networks. *IEEE Journal of Biomedical and Health Informatics* 23 (5):1911-1919. doi:10.1109/JBHI.2018.2874033
  13. Ali Z, Muhammad G, Alhamid MF (2017) An Automatic Health Monitoring System for Patients Suffering From Voice Complications in Smart Cities. *IEEE Access* 5:3900-3908. doi:10.1109/ACCESS.2017.2680467
  14. Razzak MI, Imran M, Xu G (2020) Big data analytics for preventive medicine. *Neural Computing and Applications* 32 (9):4417-4451. doi:10.1007/s00521-019-04095-y
  15. Hossain MS, Muhammad G, Alamri A (2019) Smart healthcare monitoring: a voice pathology detection paradigm for smart cities. *Multimedia Systems* 25 (5):565-575. doi:10.1007/s00530-017-0561-x
  16. Arias-Londono JD, Gomez-Garcia JA, Godino JI (2019) Multimodal and multi-output deep learning architectures for the automatic assessment of voice quality using the GRB scale. *IEEE Journal of Selected Topics in Signal Processing*:1-1. doi:10.1109/JSTSP.2019.2956410
  17. The American Heritage® Stedman's Medical Dictionary Retrieved MAy 1, 2018 from Dictionary.com website <http://dictionary.reference.com/browse/dysphonia>.
  18. Mau T (2010) Diagnostic evaluation and management of hoarseness. *The Medical clinics of North America* 94 (5):945-960. doi:10.1016/j.mcna.2010.05.010
  19. Roy N, Merrill RM, Gray SD, Smith EM (2005) Voice Disorders in the General Population: Prevalence, Risk Factors, and Occupational Impact. *The Laryngoscope* 115 (11):1988-1995. doi:10.1097/01.mlg.0000179174.32345.41
  20. Roy N, Merrill RM, Thibeault S, Parsa RA, Gray SD, Smith EM (2004) Prevalence of voice disorders in teachers and the general population. *J Speech Lang*

- Hear Res 47 (2):281-293. doi:10.1044/1092-4388(2004/023)
21. Quick Statistics: Voice, Speech, and Language. National Institute on Deafness and Other Communication Disorders. <http://www.nidcd.nih.gov/health/statistics/vsl/Pages/stats.aspx>. Accessed 1 May 2018 2018
  22. Yan Q, Yang R, Huang J Copy-move detection of audio recording with pitch similarity. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 19-24 April 2015 2015. pp 1782-1786. doi:10.1109/ICASSP.2015.7178277
  23. Eddins DA, Anand S, Lang A, Shrivastav R (2020) Developing Clinically Relevant Scales of Breathiness and Rough Voice Quality. *Journal of Voice*. doi:<https://doi.org/10.1016/j.jvoice.2019.12.021>
  24. Nemr K, Simoes-Zenari M, Cordeiro GF, Tsuji D, Ogawa AI, Ubrig MT, Menezes MH (2012) GRBAS and Cape-V scales: high reliability and consensus when applied at different times. *Journal of voice : official journal of the Voice Foundation* 26 (6):812.e817-822. doi:10.1016/j.jvoice.2012.03.005
  25. Thiruvananthapuram T, Ambikairajah E, Epps J, Enzinger E A comparison of single-stage and two-stage modelling approaches for automatic forensic speaker recognition. In: 2013 IEEE 8th International Conference on Industrial and Information Systems, 17-20 Dec. 2013 2013. pp 433-438. doi:10.1109/ICIInfS.2013.6732023
  26. Uloza V, Vegiene A, Saferis V (2015) Correlation between the quantitative video laryngostroboscopic measurements and parameters of multidimensional voice assessment. *Biomedical Signal Processing and Control* 17:3-10. doi:<http://dx.doi.org/10.1016/j.bspc.2014.10.006>
  27. Poburka BJ (1999) A new stroboscopy rating form. *Journal of Voice* 13 (3):403-413. doi:10.1016/S0892-1997(99)80045-9
  28. Rosen CA (2005) Stroboscopy as a Research Instrument: Development of a Perceptual Evaluation Tool. *The Laryngoscope* 115 (3):423-428. doi:10.1097/01.mlg.0000157830.38627.85
  29. Deguchi S, Ishimaru Y, Washio S (2007) Preliminary Evaluation of Stroboscopy System Using Multiple Light Sources for Observation of Pathological Vocal Fold Oscillatory Pattern. *Annals of Otolaryngology, Rhinology & Laryngology* 116 (9):687-694. doi:10.1177/000348940711600911
  30. Speyer R, Wieneke GH, Kersing W, Dejonckere PH (2005) Accuracy of Measurements on Digital Videostroboscopic Images of the Vocal Folds. *Annals of Otolaryngology, Rhinology & Laryngology* 114 (6):443-450. doi:10.1177/000348940511400606
  31. Patel R, Dailey S, Bless D (2008) Comparison of High-Speed Digital Imaging with Stroboscopy for Laryngeal Imaging of Glottal Disorders. *Annals of Otolaryngology, Rhinology & Laryngology* 117 (6):413-424. doi:10.1177/000348940811700603
  32. Bohr C, Kraeck A, Eysoldt U, Ziethe A, Döllinger M (2013) Quantitative analysis of organic vocal fold pathologies in females by high-speed endoscopy. *The Laryngoscope* 123 (7):1686-1693. doi:10.1002/lary.23783
  33. Manfredi C, Bocchi L, Cantarella G, Peretti G (2012) Videokymographic image processing: Objective parameters and user-friendly interface. *Biomedical Signal Processing and Control* 7 (2):192-201. doi:<http://dx.doi.org/10.1016/j.bspc.2011.02.007>
  34. Krausert CR, Olszewski AE, Taylor LN, McMurray JS, Dailey SH, Jiang JJ (2011) Mucosal Wave Measurement and Visualization Techniques. *Journal of Voice* 25 (4):395-405. doi:<http://dx.doi.org/10.1016/j.jvoice.2010.02.001>
  35. Švec JG, Schutte HK (2012) Kymographic imaging of laryngeal vibrations. *Current opinion in otolaryngology & head and neck surgery* 20 (6):458-465. doi:10.1097/MOO.0b013e3283581feb
  36. Woo P (2014) Objective Measures of Laryngeal Imaging: What Have We Learned Since Dr. Paul Moore. *Journal of Voice* 28 (1):69-81. doi:10.1016/j.jvoice.2013.02.001
  37. Muhammad G, Mesallam TA, Malki KH, Farahat M, Alsulaiman M, Bukhari M (2011) Formant analysis in dysphonic patients and automatic Arabic digit speech recognition. *Biomedical engineering online* 10 (41):1-12. doi:10.1186/1475-925X-10-41
  38. Kay Elemetric Corp. (1993) Muti-Dimensional Voice Program (MDVP) Ver. 3.3. Lincoln Park, NJ
  39. Milenkovic P, Read C (1992) CSpeech version 4 user's manual. Madison, WI
  40. Boersma P, Weenink D (2001) Praat, a system for doing phonetics by computer. *Glott International* 5:341-345
  41. Arjmandi MK, Pooyan M, Mikaili M, Vali M, Moqarehzadeh A (2011) Identification of voice disorders using long-time features and support vector machine with different feature reduction methods. *Journal of voice : official journal of the Voice Foundation* 25 (6):e275-289. doi:10.1016/j.jvoice.2010.08.003

42. Massachusetts Eye & Ear Infirmary Voice & Speech LAB (1994) Disordered Voice Database Model 4337 (Ver. 1.03) Kay Elemetrics Corp., Lincoln Park, NJ
43. Peppard RC, Bless DM, Milenkovic P (1988) Comparison of young adult singers and nonsingers with vocal nodules. *Journal of Voice* 2 (3):250-260. doi:[http://dx.doi.org/10.1016/S0892-1997\(88\)80083-3](http://dx.doi.org/10.1016/S0892-1997(88)80083-3)
44. Lin E, Jiang J, Hanson DG (1998) Glottographic signal perturbation in biomechanically different types of dysphonia. *Laryngoscope* 108 (1 Pt 1):18-25
45. Rosen CA, Lombard LE, Murry T (2000) Acoustic, Aerodynamic, and Videostroboscopic Features of Bilateral Vocal Fold Lesions. *Annals of Otolaryngology & Laryngology* 109 (9):823-828. doi:[10.1177/000348940010900907](https://doi.org/10.1177/000348940010900907)
46. Al-nasheri A, Muhammad G, Alsulaiman M, Ali Z, Mesallam TA, Farahat M, Malki KH, Bencherif MA (2017) An Investigation of Multidimensional Voice Program Parameters in Three Different Databases for Voice Pathology Detection and Classification. *Journal of Voice* 31 (1):113.e119-113.e118. doi:<https://doi.org/10.1016/j.jvoice.2016.03.019>
47. Ali Z, Elamvazuthi I, Alsulaiman M, Muhammad G (2016) Detection of Voice Pathology using Fractal Dimension in a Multiresolution Analysis of Normal and Disordered Speech Signals. *Journal of Medical Systems* 40 (1):20. doi:[10.1007/s10916-015-0392-2](https://doi.org/10.1007/s10916-015-0392-2)
48. Ali Z, Talha M, Alsulaiman M (2017) A Practical Approach: Design and Implementation of a Healthcare Software for Screening of Dysphonic Patients. *IEEE Access* 5:5844-5857. doi:[10.1109/ACCESS.2017.2693282](https://doi.org/10.1109/ACCESS.2017.2693282)
49. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE transactions on pattern analysis and machine intelligence* 24 (7):971-987. doi:[10.1109/TPAMI.2002.1017623](https://doi.org/10.1109/TPAMI.2002.1017623)
50. Lopez PG, Montresor A, Epema D, Datta A, Higashino T, Iamnitchi A, Barcellos M, Felber P, Riviere E (2015) Edge-centric Computing: Vision and Challenges. *SIGCOMM Comput Commun Rev* 45 (5):37-42. doi:[10.1145/2831347.2831354](https://doi.org/10.1145/2831347.2831354)
51. Ali Z, Elamvazuthi I, Alsulaiman M, Muhammad G (2016) Detection of Voice Pathology using Fractal Dimension in a Multiresolution Analysis of Normal and Disordered Speech Signals. *J Med Syst* 40 (1):1-10
52. Ali Z, Imran M, Alsulaiman M, Zia T, Shoaib M (2018) A zero-watermarking algorithm for privacy protection in biomedical signals. *Future Generation Computer Systems* 82:290-303. doi:<https://doi.org/10.1016/j.future.2017.12.007>
53. Markaki M, Stylianou Y (2011) Voice Pathology Detection and Discrimination Based on Modulation Spectral Features. *IEEE Transactions on Audio, Speech, and Language Processing*, 19 (7):1938-1948. doi:[10.1109/tasl.2010.2104141](https://doi.org/10.1109/tasl.2010.2104141)
54. Muhammad G, Ali Z, Alsulaiman M, Almutib K Vocal Fold Disorder Detection by applying LBP Operator on Dysphonic Speech Signal. In: Kijima H (ed) 2nd International Conference on Intelligent Control, Modelling and Systems Engineering, Cambridge, MA, USA, January 29-31 2014. WSEAS Press, pp 222-228
55. Marwan N, Carmen Romano M, Thiel M, Kurths J (2007) Recurrence plots for the analysis of complex systems. *Physics Reports* 438 (5-6):237-329. doi:<http://dx.doi.org/10.1016/j.physrep.2006.11.001>
56. Titze I (1995) Workshop on acoustic voice analysis: Summary statement. National Center for Voice and Speech, Denver, CO.,