Review Article

Nuha Qais Abdulmajeed, Belal Al-Khateeb*, and Mazin Abed Mohammed

A review on voice pathology: Taxonomy, diagnosis, medical procedures and detection techniques, open challenges, limitations, and recommendations for future directions

https://doi.org/10.1515/jisys-2022-0058
received March 08, 2022; accepted May 05, 2022

Abstract: Speech is a primary means of human communication and one of the most basic features of human conduct. Voice is an important part of its subsystems. A speech disorder is a condition that affects the ability of a person to speak normally, which occasionally results in voice impairment with psychological and emotional consequences. Early detection of voice problems is a crucial factor. Computer-based procedures are less costly and easier to administer for such purposes than traditional methods. This study highlights the following issues: recent studies, methods of voice pathology detection, machine learning and deep learning (DL) methods used in data classification, main datasets utilized, and the role of Internet of things (IoT) systems employed in voice pathology diagnosis. Moreover, this study presents different applications, open challenges, and recommendations for future directions of IoT systems and artificial intelligence (AI) approaches in the voice pathology diagnosis. Finally, this study highlights some limitations of voice pathology datasets in comparison with the role of IoT in the healthcare sector, which shows the urgent need to provide efficient approaches and easy and ideal medical diagnostic procedures and treatments of disease identification for doctors and patients. This review covered voice pathology taxonomy, detection techniques, open challenges, limitations, and recommendations for future directions to provide a clear background for doctors and patients. Standard databases, including the Massachusetts Eye and Ear Infirmary, Saarbruecken Voice Database, and the Arabic Voice Pathology Database, were used in most articles reviewed in this article. The classes, features, and main purpose for voice pathology identification are also highlighted. This study focuses on the extraction of voice pathology features, especially speech analysis, extends feature vectors comprising static and dynamic features, and converts these extended feature vectors into solid vectors before passing them to the recognizer.

Keywords: voice pathology, deep learning algorithms, classification, voice pathology detection, voice pathology databases, IoT systems, feature extraction, machine learning

*** Corresponding author: Belal Al-Khateeb,** Computer Science Department, College of Computer Science and Information Technology, University of Anbar, 31001, Ramadi, Anbar, Iraq, e-mail: belal-alkhateeb@uoanbar.edu.iq
Nuha Qais Abdulmajeed: Computer Science Department, College of Computer Science and Information Technology, University of Anbar, 31001, Ramadi, Anbar, Iraq, e-mail: nuh20c1014@uoanbar.edu.iq
Mazin Abed Mohammed: Computer Science Department, College of Computer Science and Information Technology, University of Anbar, 31001, Ramadi, Anbar, Iraq, e-mail: mazinalshujeary@uoanbar.edu.iq

1 Introduction

Normal voice is the result of the interaction of the larynx with the pulmonary air pulses, which sets the actual movement of the vocal fold toward the midline and results in the production of sounds, either aperiodic or periodic. Voice disorder is defined as any sort of abnormality that deviates from the characteristics such as "loudness, quality, pitch and/or vocal flexibility," compared with voices from the same age group, sex, and social group [1,2]. Twenty percent of young adults and children aged 3–21 years had voice problems based on a newly published survey from the National Center for Education Statistics. The American Speech-Language-Hearing Association found that voice impairment is caused by a defect in the human voice generation system [3].

Invasive surgical treatments are frequently performed to diagnose vocal abnormalities. Doctors put a probe into the mouth during endoscopic procedures, such as laryngoscopy, laryngeal electromyography, and stroboscopy. These operations are painful and can traumatize patients. Finding alternatives to these surgical treatments has been the subject of extensive research. One of these alternatives is utilizing voice signal processing to detect vocal pathology [4]. Extraction and analysis of voice features are required for vocal pathology detection using voice signals. Voice samples were taken in a specific environment. The extracted voice features from the voice signals were then examined. Voice samples were then divided into two groups: normal and abnormal. The right technology must be chosen carefully considering voice signal identification. The classification algorithm is the most widely used technique to identify and differentiate between two or more subjects [5]. Machine learning has shown efficient performance in various medical domains, including cancer diagnosis and classification [6], 3D brain reconstruction [7,8], and Alzheimer's disease [9–11].

The previous research has shown that the accuracy level of a classifier is significantly dependent on this measurement. However, numerous concerns and challenges are associated with pathology detection approaches based on voice signals. Several issues such as (1) selection of adequate voice features and (2) identification of an appropriate classifier [5] are also important. The current study aims to present and discuss an exhaustive analysis of various voice abnormalities, machine learning, voice disorder databases, classifiers, and feature extraction methods. Moreover, this study has been employed for the development of a mechanism that works automatically for the classification and detection of voice pathologies. The primary contributions of this work are presented as follows.

- This review covered voice pathology taxonomy, diagnosis, medical procedures and detection techniques, open challenges, limitations, and recommendations for future directions to provide clear background for patients.
- This study presents medical conditions for voice pathology with their categories, and this includes swallowing difficulties, speech defects, language defects, cognitive communication defects, and social communication defects.
- Standard databases were used in most articles that we have reviewed in this article, including Massachusetts Eye and Ear Infirmary (MEEI), Saarbruecken Voice Database (SVD), and Arabic Voice Pathology Database (AVPD), and also, highlighted classes, features, and main purpose for voice pathology identification.
- This study focuses on voice pathology feature extraction, especially the speech analysis, extended feature vector made of static and dynamic features, and converts these extended feature vectors into more solid vectors before passing them on to the recognizer.
- The current study demonstrates the importance of AI approaches, such as machine and deep learning, which are used to classify the cases into healthy or pathological.
- This study also demonstrates that Internet of things (IoT) is frequently only partially and correctly utilized in the healthcare domain due to a significant gap between the data used and the ability to process and analyze the approach. The largest challenge for each patient, particularly those living in rural places, lies in their inability to get doctors and remedies in emergencies. Therefore, an innovative system based on the IoT is required in these areas.

The following is a breakdown of the structure of this article. Section *2* demonstrates an example of voice pathologies. Section 3 explains the voice pathology database. Section 4 discusses approaches for extracting voice pathology features. Section 5 presents voice pathology classification algorithms. Section 6 explains IOT voice pathology. Section *8* concludes this article.

2 Voice disorders

This section summarizes the voice disorders as follows:

2.1 Laryngitis

The swelling of the vocal cords is known as laryngitis, resulting in a hoarse voice [12].

2.2 CYST

A development is observed under the superficial layer of the vocal fold mucosa. If a void occurs between the layers, then the vocal folds fail to vibrate appropriately. A cyst can also thicken the mucosa of the vocal folds, contributing to the impossible normal vibration [13].

2.3 Polyp of the vocal fold

This polyp arises from the mucosa of the vocal folds. If these polyps are strong or fluid packed, then they can expand to remarkably large sizes. The vibratory intensity of the vocal folds is determined by their size and position [13].

2.4 Nodules

A vocal fold nodule is a symmetrical prose located on both sides of the vocal folds in the middle of the voice box. Dysphonia is the most common vocal symptom when nodules in the vocal folds prevent the entire closing of folds [13].

2.5 Paralysis

This phenomenon is the failure of the vocal cords to move. One or both vocal folds have become immovable, leading to a substantial space between them. This gap allows air to escape but obstructs the natural movement [13].

2.6 Sulcus

This disorder is a linear recession that runs parallel to the free border on the mucosal surface of the vocal folds under varied depth and bilateral symmetry. The vocal folds are prevented from moving properly due to the sulcus vocalist, resulting in mild to severe dysphonia [13].

3 Voice pathology database

The voice pathology database is a critical component of the automatic voice disorder detection (AVDD) system. Samples of healthy and disordered voices are included in the dataset. These samples may include sustained vowel phonation or continuous speech. Several common databases such as the MEEI, the SVD, and the AVPD are used in most studies [14].

3.1 Massachusetts eye and ear infirmary

A database was created by the MEEI lab for voice and speech. It contains around 1,400 voiced, sustained vowel/a/samples, as well as the initial half of the Rainbow Passage. Kay Elemetrics is the company that sells this database [15,16]; it was captured in two different settings. Normal samples were experimented with at 50 kHz, while pathological samples were tested at 25 or 50 kHz. A spreadsheet with clinical and personal information from the individuals as well as the findings of the acoustic analysis of the recordings collected from Kay's MultiDimensional Voice Program (MDVP) are also included in the database. The Computerized Speech Lab (CSL) of Kay was also used to make the recordings under similar acoustic settings. Each participant was instructed to generate a sustained phonation of the vowel/ah/ at a comfortable pitch and loudness for at least 3 s. The technique was performed three times for each participant, and the best sample was chosen for the database by a speech pathologist [17]. Despite its popularity, this database has several drawbacks. The database is used in most investigations of vocal pathology identification and classification and utilizes different settings and frequencies of voice samples to capture healthy and disordered voices (Kay Elemetrics confirmed this data). A variety of vocal disorders could also be detected by analyzing changes in the voice muscles that can activate and increase voice efficiency [15] (Figure 1).

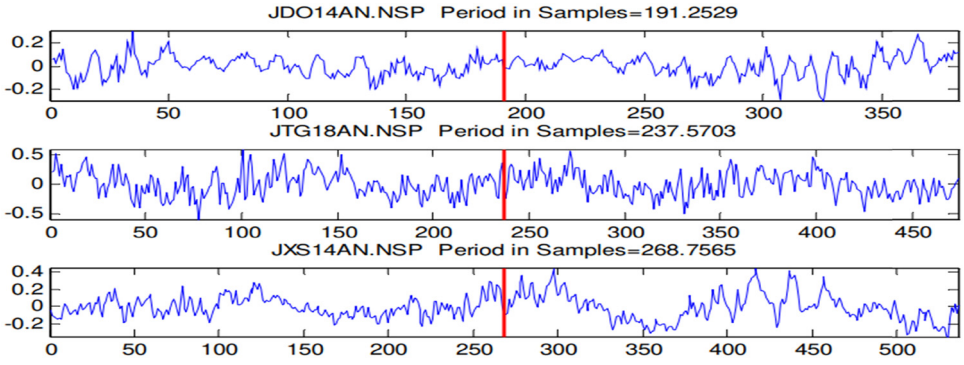

Figure 1: Three types of signal examples from the MEEI database [15].

3.2 Saarbruecken voice database

SVD is an available free database [18] maintained by the Institute of Phonetics at Saarland University. This database includes/a/,/i/, and/u/sustained vowels with several intonations, including low-high-low, low, normal, and high, as well as a spoken sentence in German, "Guten Morgen, wie geht es Ihnen?," which translates to "Good morning, how are you?" These characteristics contribute to the remarkability of this database. The recorded voices of the SVD database were all sampled at 50 kHz with a 16-bit resolution. This database is also new, and only a small number of studies have used it in the detection of voice disorders [15]. Files can be downloaded from the website provided in [18]. The voice and EGG signals are saved for each recording. Both signals may be exported in their native file formats (NSP and EGG) as well as WAV,

thus creating a zip file on the server containing the chosen files. The procedure might take several minutes. Finally, a link to download the produced zip file appears [18] (Figure 2).

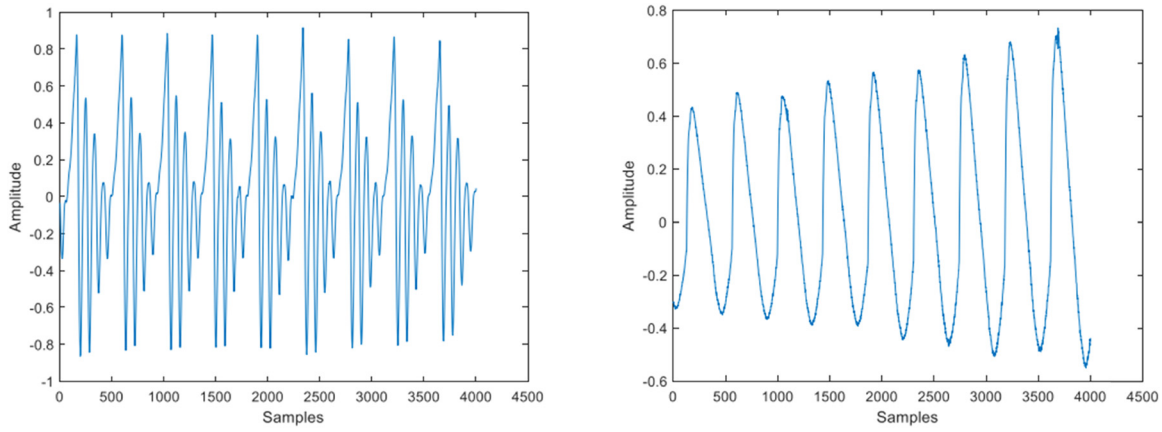

Figure 2: Voice and EGG signal examples from the SVD database [13].

3.3 Arabic voice pathology database

Samples of this database were recorded at the Communication and Swallowing Disorders Unit (King Abdul Aziz University Hospital, Riyadh, Saudi Arabia) during several sessions in a sound-treated environment by skilled phoneticians using a defined recording methodology [15,19]. The gathering of the database was one of the key responsibilities of a 2-year initiative supported by the National Plans for Science and Technology (NPST) in Saudi Arabia. Database protocol was created in such a way that would avoid the flaws of the MEEI database [15,20]. The database contains various samples, including speeches from individuals with vocal-fold diseases, sustained vowels, as well as recordings from those with normal speech. Healthy and ill vocal folds were recognized after a clinical examination with laryngeal stroboscopy. On a scale of 1–3 severity, the frequency of voice problems was graded in cases of pathology, with 3 representing the most severe case. The panel of three certified medical experts agreed on a severity rating for each sample. Three types of text are found in the recording: (a) continuous speech, (b) three sustained vowels with the onset and offset information, and (c) isolated words, numbers, and common words in Arabic [15].

The text was carefully chosen to include all the phonemes of Arabic. All speakers recorded three utterances of each vowel (/a/,u/, and/i/) to prevent burdening patients, but isolated words and continuous speech were only recorded once. The speech was recorded using the CSL application with a sampling frequency of 44 kHz in the database. Different expert specialists from King Abdul Aziz University Hospital analyzed and validated the vocal problems listed in this database [15]. The AVPD contains 366 samples of normal and diseased patients. Normal participants account for 51% of the total subjects, with the remainder divided among five voice abnormalities (sulcus, nodules, cysts, paralysis, and polyps); 40% are female, while 60% of the subjects are male. All texts were captured and saved in two separate audio formats, namely, WAV and NSP, including three vowels with a repeat, namely, Arabic numbers, Al-Fateha, and common terms. The recorded samples were separated into 22 segments: 6 for vowels (three vowels plus their repetition), 11 for Arabic numerals (zero to ten), and 2 for Al-Fateha (divided in such a way that the first half may be used to train the system and the second part can be used to test the system), and 3 segments for the common word [21].

Many studies use different datasets and classification methods for the detection of various voice pathologies as presented in Table 1 (Figure 3).

Table 1: Similar voice pathology detection and classification method studies

Ref.	Classifier	Dataset	Pathologies
[15]	Kay Pentax CSL Model	MEEI SVD AVPD	(1) Unilateral vocal fold paralysis (2) Cyst of vocal fold (3) Polyp on the vocal folds
[21]	The AVPD has five forms of vocal fold diseases; thus, five classes were presented in this study	AVPD MEEI	(1) Sulcus (2) Nodules (3) Cysts (4) Paralysis (5) Polyps
[22]	GMM	SVD	(1) Vocal polyp (2) Sulcus (3) Vocal nodules (4) Vocal cyst (5) Vocal paralysis (unilateral or bilateral)
[23]	SVM	MEEI SVD AVPD	(1) Cysts in the vocal folds (2) Vocal fold paralysis on one side (3) Polyp of the vocal fold
[24]	CNN	SVD MEEI	(1) Vocal cyst (2) Vocal polyp (3) Vocal paralysis (unilateral)
[25]	SVM	MEEI SVD AVPD	(1) Cysts in the vocal folds (2) Unilateral paralysis of the vocal folds (3) Polyps on the vocal folds
[26]	GMM	MEEI	(1) Vocal polyp (2) Keratosis (3) Vocal paralysis (4) Nodules in the vocal fold (5) Dysphonia
[27]	CNN	SVD	Dysphonia
[28]	CNN	SVD MEEI	(1) Paralysis (unilateral) (2) Polyp (3) Cyst of vocal folds
[29]	SVM	SVD	(1) Dysphonia (2) Laryngitis (3) Recurrent sparse
[30]	DPM	SVD	(1) Reinke's edema (2) Laryngitis
[31]	MFCC	SVD MEEI AVPD PDA	(AVPD) Paralysis, sulcus, polyp, cyst, nodules (MEEI) Gastric reflux, hyperfunction, AP squeezing, ventricular compression, and paralysis (PDA) nodule bilateral, edema pure bilateral, polyp, sulcus (SVD) Hyperfunctional dysphonia, recurrent sparse, laryngitis, functional dysphonia, dysphonia
[32]	CNN, RNN	SVD	(1) Cysts, (2) polyps, (3) nodules, (4) paralysis, (5) sulcus
[33]	OSELM	SVD	(1) Cyst (2) Polyp (3) Paralysis
[13]	CNN: ResNet50, Xception, and MobileNet	SVD	(1) Cysts, (2) polyps, (3) nodules, (4) paralysis, (5) sulcus

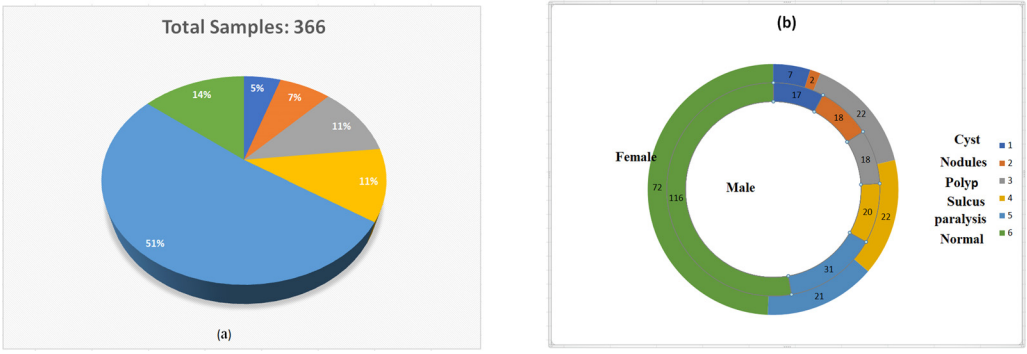

Figure 3: (a) AVPD distribution of healthy and voice-disordered people. (b) The number of male and female samples for each disease and normal subjects [21].

4 Feature extraction techniques of voice pathology

Feature extraction is the calculation of a sequence of feature vectors that provides a compact representation of the input speech signal. This method is generally divided into three components. The first stage is the speech analysis or acoustic front-end, which analyzes the speech signal spectra temporally and creates raw characteristics reflecting the power spectrum envelope of brief speech intervals. The second stage involves the generation of a feature vector that comprises static and dynamic information. The final stage compresses and strengthens the extended feature vectors before handing them to the recognizer [34].

4.1 Mel frequency cepstral coefficients

Mel frequency cepstral coefficients (MFCCs) are the most often used characteristics in the detection of voice pathologies. Using MFCCs to define the human voice generation system is widely accepted. MFCC is a useful method for detecting vocal impairment [5]. MFCC is one of the most commonly used feature extraction methods in speech recognition. MFCC depends on the Mel scale, which is based on the human ear scale; thus, features of the frequency domain, including MFCCs, are more precise than those of the time domain [35,36]. MFCC, which is produced from the fast Fourier transform signal, represents the true cepstral of a windowed short-time signal (FFT). A nonlinear frequency scale is used to simulate the auditory system activity to distinguish MFCC from the true cepstral. Furthermore, these coefficients are stable and reliable considering fluctuations in speakers and recording settings. Simultaneously, MFCC is a technique used to extract features by determining parameters from samples of speech that are similar to those used by humans for hearing speech and simultaneously downplaying all other data [34]. As illustrated in Figure 4, MFCC for various speech samples offers an advantage over other voice characteristics because they can characterize the geometry of the vocal tract completely. An exact representation of the phoneme generated by the vocal tract may be calculated after accurately defining the vocal tract. The envelope of a short-time power spectrum depicts the curvature of the vocal tract, and MFCCs record this envelope properly [5].

4.2 Spectrograms

A voice waveform comprises a succession of changing events. This temporal variation relates to spectral properties that drastically shift over time. An STFT is used instead because a single Fourier transform cannot capture this kind of rapid time-varying signal [37]. The STFT, when seen through a sliding window, comprises a separate Fourier transform for each segment of the waveform. The spectrogram can be presented in three dimensions to demonstrate the distribution of power densities with time and frequency, as

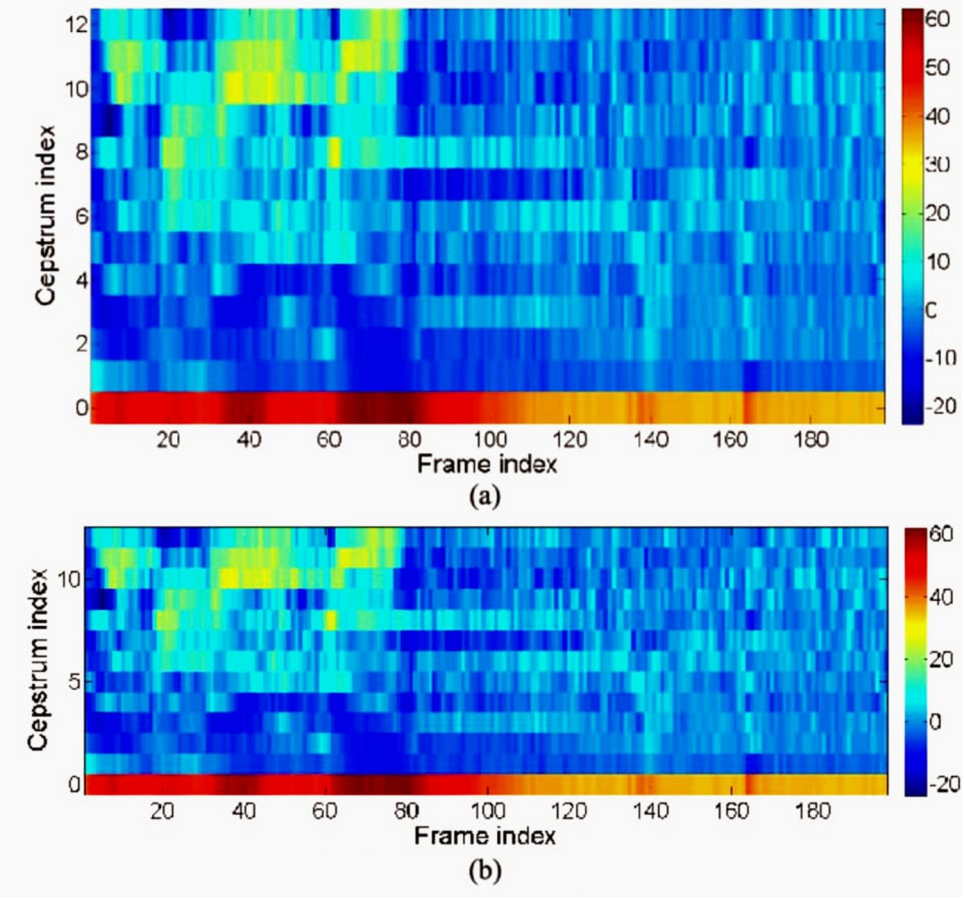

Figure 4: The MFCC of normal and pathological voice samples [5].

shown in Figure 5. The figure reveals that the power density distribution of the voice signal substantially changes with time and frequency and may be used to distinguish between normal and sick voices. The graph also demonstrates that the power distribution for normal voice is stable with time and frequency. However, the results for unusual voices are different. The spectrogram is regarded as an effective signal for differentiating between sick and normal voices [5].

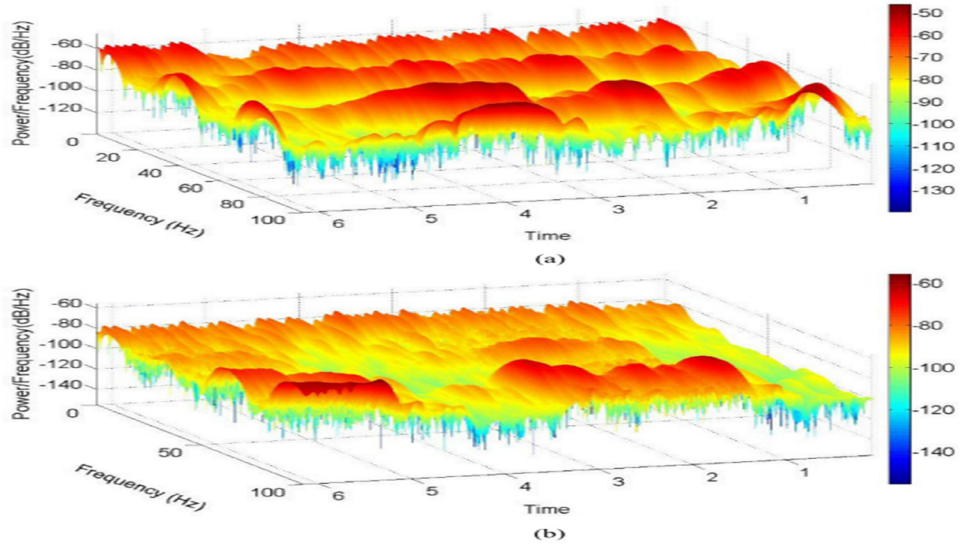

Figure 5: The voice spectrograms of normal and diseased voices [5].

4.3 Temporal and spectral features

Time-frequency analysis is accomplished by using signal energy, entropy, and zero-crossing rate, as well as a discriminative paraconsistent machine with a 95% accuracy rate, resulting in the SVD dataset in refs [1,38]. Aperiodic and periodic features in noise signals can be considered in vocal fold pathology detection in remarkably advanced stages. By contrast, the sight of the first spike in the envelope enclosing the spectrum can be utilized at slightly reduced complexity levels [1,39]. Smartphones helped to detect any sort of abnormalities in the speech of people with an increased risk of contracting PD [1,40]. Biomarkers based on voice-based mechanisms acquired via cellphones were found to improve the diagnosis approach in neurodegenerative disorders as well as provide important enhancements in stratification for future neuroprotective therapeutics for Parkinson's disease. A perceptual evaluation technique for dealing with anomalies was proposed and applied to a few sets of decisions that could be handled only by an autonomous system [1,41], thus proving its relevance in anomaly detection. The human auditory mechanism-based intelligent system, which can detect and classify numerous classes of disorders related to vocal folds, was also introduced [1,26]. A voice disorder detection system to identify any abnormalities by examining the signal originating from the speech through a linear prediction analysis was proposed; this system can classify and distinguish the features between normal and abnormal samples [1,42].

4.4 Online sequential extreme learning machine

A method of offline supervised batch learning, which requires the availability of all data samples to complete the training process, is called ELM [33,43,44]. However, in a real-world application, packets of data are gathered over time; that is, only some data are available at once. Consequently, an upgraded ELM version, dubbed Online sequential extreme learning machine (OSELM), was presented in refs [33,45] to address the aforementioned problem. This algorithm is designed to address new requirements in a variety of online learning applications. OSELM is a quick method that is preferable among algorithms because it decreases the need for retraining upon receiving new data. From training data, OSELM may learn by using a chunk-by-chunk technique with a fixed or variable number of chunks. The OSELM algorithm comprises the following three layers: the hidden layer, the output layer, and the input layer. The input layer is where the retrieved features are placed. Biases are found in the hidden node, while the final classifications of the algorithm are in the output layer [33].

4.5 Body linear predictive codes

Signal compression is desirable for effective transmission and storage. Digital signals are compressed before transmission to maximize channel use on wireless media. The most common medium or low bit-rate coder is linear predictive codes (LPCs) [34,46]. The power of the signal spectrum is calculated by LPC and employed for formant analysis [34,47]. LPC is also a well-known formant estimation method. This method is one of the most powerful techniques in the speech analysis [34,48]. An output of residual error is produced when a voice signal is processed by a speech analysis filter to reduce redundancy. In comparison to the original signal, it may be quantized with a reduced number of bits. Therefore, rather than sending the full signal, the residual error and speech parameters can be provided, allowing the reproduction of the original signal. The least mean square error theory is used to create a parametric model, which is referred to as linear prediction. The speech signal is approximated as a linear combination of its previous samples by using the aforementioned approach. The formants are characterized in this method by the obtained LPC coefficients. The formant frequencies are the frequencies where the resonant peaks emerge [34,49].

4.6 Perceptual linear prediction

The perceptual linear prediction (PLP) model was created by Hermansky; it is a human speaking model based on the psychophysics of hearing [34,50,51]. PLP eliminates irrelevant speech information, resulting in a high classification rate. Theoretically, PLP and LPC are identical, but spectrum properties must be modified to match the human auditory system. The PLP speech analysis approach is better suited to human hearing than linear prediction coding (LPC). The fundamental distinction between PLP and LPC analysis techniques is that the LP model implies that the vocal tract has an all-pole transfer function with a specific number of resonances inside the analysis band. This notion is false because hearing spectral resolution drops with frequency beyond 800 Hz; hearing is also sensitive in the middle frequency region of the audible spectrum [34,51].

4.7 Fisher discriminant ratio

Fisher discriminant ratio (FDR) may be used to differentiate across groups and focus on a specific group; it has also been frequently employed in the categorization of HSI for the identification of closely related features. The disjoint subsets (views) were constructed by decomposing the features into disjoint subsets (views) considering that each view is sufficient to produce a robust classification model [52,53]. Learning is conducted using complementing knowledge from several viewpoints [54]. Information is repetitive between bands, but the generation of appropriate views allows the system to learn from the data provided by multiview. These viewpoints are used to construct ensembles to determine where the ensembles vary [52].

5 Voice pathology classification algorithms

One of the most important tasks in voice signal analysis is classifying a specific signal into one of a few predefined categories to arrive at a diagnostic judgment on the vocal problem. The classification of a particular voice signal is quite useful in the diagnostic process. Several classifiers have been used to detect vocal impairment [5]. The primary goal is to distinguish between normal and abnormal speech samples, and the procedure is broken down into three steps: feature extraction, feature selection, and feature classification using a classifier. The audio samples are selected from a voice database, and the required characteristics are extracted. The chosen characteristics are then optimized using an algorithm and sent to the classifier. The classifier divides the speech samples into normal and abnormal categories. Classifiers, such as k-NN and SVM, are commonly used to detect voice dysfluencies. Neural networks and the hidden Markov model (HMM) are among the most popular nonlinear classifiers. SVMs are commonly utilized because they attempt to construct an ideal wall in the feature space that divides the two classes. Two judgments are involved in the two classes. The SVM maximizes the difference in points between the two classes. Linear separation of data is difficult in practice; in such circumstances, SVMs employ the kernel technique to translate the data into a high-order dimensional space and create the dividing wall (hyperplane) in the broad area [1] (Table 2).

6 IoT-based voice pathology

The IoT provides a framework for sensors and devices to connect smoothly inside a smart environment and a simple way to transmit information across platforms. IoT is positioned to become the next breakthrough technology with the recent adaptation of various wireless technologies, thereby maximizing the

Table 2: Examples of characteristics and classifiers used in similar studies explain the most commonly used classifiers

Authors (year)	Features extracted and classifiers	Dataset	Remarks	Best accuracy
[15] Al-Nasheri et al. (2017)	MDVP parameters, FDR, Kay Pentax CSL Model	MEEI	1. The best results are obtained using SVD features because the diseased samples are extremely severe, demonstrating distinctions between healthy and disordered samples	99.68%
		SVD	2. The best accuracy was 99.68% when utilizing the top three parameters from the SVD database	
		AVPD	3. vAm (peak amplitude variation period-to-period), APQ (amplitude perturbation quotient), and PFR (phonatory fundamental frequency) were the top three characteristics that provided the SVD outstanding accuracy	
			4. aVm is more capable of distinguishing between normal and diseased voices than the two other parameters and has a larger capacity than the two other factors to distinguish abnormal samples	
			5. One of the limitations lies in the absence of an exact relationship between the numerical parameters of the acoustic analysis and the auditory-perceptual qualities of voice	
			6. Regardless of the database employed, developing strong features that can consistently distinguish between normal and sick samples is necessary	
[21] Mesallam et al. (2017)	MFCC, PLP, LPCC, LPC, RASTAPLP, MDVP	AVPD	1. The AVPD has a healthy and ill patient population. Determining the number of effectively detected normal and diseased samples by the system is impossible when data are unbalanced, thus resulting in a problematic interpretation	93.6%
	HMM, SVM, GMM, VQ	MEEI	2. The AVPD records all samples in a controlled environment, which is critical	
			3. The AVPD database has a detection accuracy of 79.5%, whereas the MEEI database has a detection accuracy of 93.6%	
			4. The loss of information in the signal-to-noise ratio is another disadvantage attributed to persistent phonation, which requires a complete recording, including quiet at the beginning and conclusion of the recording, to compute	
			5. Automatic methods are occasionally unable to distinguish between healthy and ill patients. Therefore, perceptual severity is included in the AVPD and is graded on a range of 1–3, with 3 indicating a severe voice problem	
			6. The balance considerably reduced the total size of the sample, affecting the reliability of the results	

(*Continued*)

Table 2: *Continued*

Authors (year)	Features extracted and classifiers	Dataset	Remarks	Best accuracy
[22] Muhammad et al. (2017)	Entropy, contrast, energy, homogeneity GMM, co-occurrence matrices	SVD	1. The voice-only system obtained 99% accuracy, while the EGG-only system achieved 89.4% accuracy 2. The system took an average of 1.45 s per patient. This period includes the time for processing and classification 3. Voice-only signal can more accurately assess voice pathology than EGG-only signal; nevertheless, the decision based on an EGG-only signal may be inaccurate for severe voice pathology	99%
[55] Hossain et al. (2017)	Employing an iterative adaptive inverse filtering approach to separate the glottal signal from the spoken signal EGG: shape and cepstral domain GMM	SVD	1. The best accuracy of 93% was achieved using 16 combinations for each input signal. Two and four mixes performed poorly, particularly for the EGG signal 2. The results demonstrate that the voice signal is more effective than the EGG signal in detecting speech pathology	93%
[26] Ali et al. (2018)	Phenomena of critical bandwidths FCB, GMM	MEEI	1. The proposed approach has 99.72% accuracy in detecting pathology 2. A paralysis versus all other illnesses experiment is also undertaken, with a 99.13% accuracy rate 3. The findings suggest that the proposed technique is accurate and dependable in assessing vocal fold disorders and may be used for remote diagnosis 4. Compared with current disorder assessment methods, the suggested system has a superior performance 5. The outcome of the FCB is superior to that of the FCBD because it is achieved by using limited features and Gaussian mixtures. Fewer coefficients and mixes of Gaussian distributions result in minimal computations and short-run times. This finding indicates that the results obtained with a large number of Gaussian mixtures are reliable	99.72%
[56] Roy et al, (2019)	MFCC, RNN, GRU, LSTM	SVD	1. They demonstrated that convolutional neural networks can provide improved accuracy considering long sequences in this study 2. Even with a complicated model like the one used here (LSTM), the vanishing gradient problem also affects the recurrent network	75.64%
[57] Ghoniem (2019)	Spectrogram segments, CNN	MEEI	1. Results show that the accuracy of speech pathology classification improved by 5.4% compared with the basic CNN when the suggested method is used	99.37%

(*Continued*)

Table 2: *Continued*

Authors (year)	Features extracted and classifiers	Dataset	Remarks	Best accuracy
			2. The ResNet34 architecture was used to achieve the best CNN-GA results. This result is due to the remarkable representational capabilities of residual networks 3. Considering classification accuracy, sensitivity, and specificity, the CNN-GA is better than the basic CNN (the CNN-SA, CNN-PSO, and CNNBA algorithms) 4. Three models of CNN were used to evaluate the performance of the CNN-GA method, and results show that the ResNet34 model was superior to others 5. Training of the CNN by using the gradient descent approach will result in solutions that are stuck in the local optima. Moreover, the performance of any trained CNN is determined by its initial weights. Therefore, the GA is employed in this study to find the best weight set among various initial weight settings	
[58] AL-dhief et al. (2020)	MFCC, OSELM	SVD	1. The results reveal that maximum accuracy, sensitivity, and specificity are 85, 87, and 87, respectively 2. The performance of the OSELM algorithm efficiently distinguishes between healthy and abnormal sounds. Compared with other deep learning algorithms, OSELM is a fast and accurate classifier. OSELM does not retrain the entire database when new data are obtained; instead, it trains and evaluates the new data	85%
[59] Narendra et al. (2020)	OpenSMILE features, SVM	SVD	1. Deep learning models trained with the glottal flow have higher classification accuracy than raw speech 2. If only a small quantity of training data are available, inclusion of phonemic and speaker-specific information complicates the deep learning network of the detection issue (which is the situation in the current investigation due to the training of systems with diseased voices) 3. However, results produced by CNNs with various convolutional layer counts were always less than those provided by the best classical pipeline systems	91.88%
[60] Tuncer et al. (2020)	MCMTTP, KNN, LD, DT, SVM, BT	SVD	1. The fused features are effective for characterizing disease detection based on voice	100%

(*Continued*)

Table 2: *Continued*

Authors (year)	Features extracted and classifiers	Dataset	Remarks	Best accuracy
			2. MCMTTP outperforms traditional TP in feature generation due to the many possibilities for center values and thresholds to produce features entirely as an inception network 3. Voice-based illness diagnosis requires numerous multiclass categorizations. Frontal resection, cordectomy, and spastic dysphonia disorders are employed in this study for performance evaluation by using these diseases 4. Limitations of NCA include the ability to choose redundant features because all characteristics have positive weights, as well as the inability to find optimal features. An iterative NCA was proposed to solve this restriction by allowing the selection of optimal and nonredundant features 5. This approach has nine levels. The center and threshold settings are modified iteratively at each stage	
[33] AL-DHIEF et al. (2021)	MFCC, OSELM	SVD	1. The OSELM algorithm is quick and accurate in the acoustic system 2. The best results achieved included 91.17% accuracy and 87.55% G-mean 3. The OSELM algorithm has difficulty classifying voices with various vowels and phrases, resulting in decreased accuracy 4. The input weights for OSELM are randomly created; thus, the accuracy is inconsistent 5. OSELM is used for testing and training with SVD only in a system of voice pathology	91.17%
[13] MUHAMMAD AND ALHUSSEIN (2021)	Bi-directional LSTM, CNN	SVD	1. The proposed approach attained an accuracy level of more than 95% 2. Bimodal inputs outperformed single inputs 3. The Xception model outperformed the two other CNN models 4. MobileNet performed admirably despite employing fewer parameters than the two other models 5. The proposed system outperformed all other compared systems 6. The system outperformed the single modality 7. The ResNet50 and MobileNet models produced lower outcomes than the Xception model of CNN 3. ResNet50 and MobileNet have a 224×224 input size, while Xception has a 299×299 input size	95%

(*Continued*)

Table 2: *Continued*

Authors (year)	Features extracted and classifiers	Dataset	Remarks	Best accuracy
[61] Lee (2021)	MFCC, LPCC, HOS, FNN, CNN	SVD	1. The combination of HOS properties in a given vowel improves classification accuracy 2. CNN classifier for mixed samples produced a considerable outcome 3. CNN and LPCC attained the best accuracy of 82.69% for the/u/ vowel in men 4. Experimenting with female or male samples with single data was more successful than that with a mixture of these samples 5. Discerning between normal and abnormal voices was possible by using a mix of parameters and deep learning approaches 6. The normalized kurtosis calculated from pathological signals is larger and more widely dispersed than that calculated from normal speech signals 7. The normalized skewness extracted from a pathological voice has a lower average value and a broader range than that extracted from a normal voice 8. Statistical differences were observed in some parameters between normal and disordered speech sounds	82.69%
[62] Fan et al. (2021)	MFCC, FC-SMOTE	MEEI SVD	1. The proposed method may substantially increase diagnostic accuracy; in real applications, FC-SMOTE is preferable for imbalanced voice diagnosis 2. The recall, specificity, G value, and F1 value of the VPD system with FC-SMOTE produced high results 3. The suggested strategy has a substantial overall influence on the number of each correct predictions based on the multiclass confusion matrix of the 10 models 4. Compared with the AUC value of multiclassification in the original class-imbalanced dataset, the VPD system with FC-SMOTE boosts the AUC values of all models. This technique enhances the AUC values of LR, NB, DT, SVM, KNN, RF, XGBoost, GBDT, MLP, and CNN in multiclassifications 5. Deep learning model outperforms other models, including MLP, according to the SVD database 6. The recognition capacity of the classifier is biased in favor of the majority class samples (pathological type), whereas the minority	100%

(*Continued*)

Table 2: *Continued*

Authors (year)	Features extracted and classifiers	Dataset	Remarks	Best accuracy
			class samples (normal type) have no recognition capability. Therefore, the overall accuracy results of the model are misleading 7. Considering time consumption for the machine learning model, a single classifier has the shortest training time, while a deep learning model has the longest 8. Pathological speech classification of an imbalanced class dataset does not generate satisfactory results in these traditional machine learning algorithms 9. The overall performance of the machine learning model when dealing with SVD is worse than that of the MEEI database 10. The experimental results reveal that without the FC-SMOTE algorithm, the VPD system performs poorly in detecting minority classes and can only multiclassify problematic speech types	
[32] Syed et al. (2021)	MFCC features, ZCR, energy entropy, energy, CNN, RNN	SVD	1. Results show that DNN is weaker than LSTM-RNN 2. Investigating classification failures is difficult because neural networks are usually "black boxes." The major cause of the error was PE and inadequate documentation due to a lack of or poor picture quality; instead, inference based on context was required. However, the generalizing model inference is still problematic 3. In addition to the training constraints given by the size of data sets, all these faults need intricate reasoning, which might limit the design of the models	87.11%

possibilities of the Internet. The adoption of IOT in smart cities has increased, with an emphasis on developing intelligent systems, such as smart workplaces, smart retail, smart agriculture, smart energy smart transportation, smart healthcare, and smart water [63–65]. A large number of communication devices are incorporated with sensor devices in the real world under the IoT paradigm, as indicated in Table 3. Data-gathering devices capture information and send it via embedded communication devices. Various communication systems, including Wi-Fi, GSM, ZigBee, and Bluetooth, are also hired in connecting continuous network devices and objects. These devices broadcast and receive data between remote devices, enabling direct integration with the physical environment to improve living standards via computer-based systems [63]. With the large volume of data and services offered by heterogeneous networks, IoT has envisioned a connected world reality [58,66].

Table 3: Examples of deep learning models and IOT samples [71]

DL model	Sample of IOT applications	Characteristics
RNN	• Identify movement pattern • Behavior detection	• Useful in IoT applications that deal with time-sensitive data • Processes data sequences through internal memory
CNN	• Detection of plant diseases • Detection of traffic signs	• The convolution layers account for most computations • Contains lesser connections than DNNs • A large training dataset is required for visual tasks
LSTM	• Human activity recognition • Mobility prediction	• Good results with data of a significant time lag • Memory cells are protected by gates
AE	• Machinery fault diagnosis • Emotion recognition	• Appropriate for feature extraction, decrease in dimensionality • Contains an equal number of input and output units • Data from the input are reconstructed by the output • Handles unlabeled data
RBM	• Indoor localization • Energy consumption prediction	• Appropriate for classification dimensionality reduction and feature extraction • Precise training method
DBN	• Classification of faults • Identification of security threats	• Suitable for discovering hierarchical features • Layer-by-layer greedy network training
VAE	• Intrusion detection • Failure detection	• A type of autoencoder. • Appropriate when labeled data are lacking
GAN	• Way finding and localization • Image to text	• Appropriate for noisy data • Comprises two networks: a discriminator and a generator
Ladder Net	• Face recognition • Authentication	• Comprises three networks: a decoder with two encoders • Suitable for noisy data

By contrast, cloud computing has risen to prominence, allowing for massive storage and data sharing [58,66]. The combination of cloud and IoT can introduce new and exciting possibilities for both technologies [58,67]. These technologies have the potential to open up a new vista of service sharing, device connectivity, pervasive sensing, as well as enhanced cooperation and communication between people and objects in a dynamic and dispersed performance environment [58,68]. Consequently, various applications are urgently needed in a variety of disciplines within the healthcare industry. Voice pathology is one of the most vital fields in the medical profession. Many people suffer from vocal disorders due to a variety of factors, including severe organ damage, smoking, air pollution, and stress. New studies indicate that more than 7.5% of the total population of the United States suffers from voice pathology. Persons in specific occupations, such as teachers and singers, are particularly affected by voice pathologies because they frequently use their voice [58,69].

Approximately 20% of American teachers have been infected with voice problems [58,69]. Two types of voice pathology detection and assessment are available: objective and subjective. If the method is right, then objective evaluation does not require any extra equipment, and the findings are always unbiased. Meanwhile, subjective evaluations necessitate specialized technology and highly qualified specialists,

resulting in a significant expense. Furthermore, subjective assessments vary in accordance with doctors and are dependent on doctor's knowledge. By contrast, the objective assessment may only be used for initial screening, and the ultimate judgment must be made by medical specialists [58,70].

7 Conclusion

The developments of IoT systems, fog computing, and deep learning methods are currently used in various ways in the healthcare sector. A comprehensive evaluation and taxonomy of the latest research and techniques have been offered in this article to leverage deep learning and IoT in a variety of healthcare applications, primarily in voice pathology identification methods. Furthermore, other issues such as the main feature extraction methods, main datasets used, major classification approaches, research outcomes, key issues, different applications, and recent state-of-the-art applied in voice pathology diagnosis have been emphasized and examined.

The employment of machine learning algorithms in health sectors is predicted to reduce healthcare costs, improve quality of life, and enhance user experience. In addition, these algorithms provide approaches, methods, and tools that may be used to diagnose various speech disorders and will generally help in the improvement of healthcare, specifically with the identification of voice pathologies.

Despite the positive outcomes obtained using machine learning and current technology in healthcare applications, several concerns and unsolved obstacles that require further exploration and discussion, such as the accurate prediction of a large number of parameters, still remain. A large volume of data should be supplied to achieve this goal. Furthermore, understanding diseases and their variants is more challenging than other tasks (e.g., identify the type of voice problem from speech). Therefore, from the standpoint of big data, having a substantial amount of medical data is critical for training strong and effective machine learning models [21]. Healthcare data are extremely complex, incomplete, and unclear. Training a good machine learning model with such a diverse and large set of data is challenging, and several factors such as data sparsity, redundancy, and missing values must be considered.

Ailments change and progress over time. Various models of machine learning are suggested in many sectors of healthcare considering static vector-based inputs. These inputs could not be handled together with the time factor. Thus, a new machine learning approach that can deal with temporal medical data will be an essential component that must be developed. Moreover, creating a new machine learning technique that considers dynamic inputs is necessary, and difficulties in healthcare and biomedicine are becoming increasingly complicated. The diseases are extremely diverse, and most of their reasons remain unknown. In addition, the number of patients in a realistic clinical trial is typically limited; thus, the authors failed to request as many additional patients as needed. A paucity of medical training data is observed for machine learning models. Healthcare organizations deal with multipatient settings where many caregivers perform their duties. In this context, proper identification of caregivers and patients is necessary. Protecting the data gathered by a range of devices and sensors in healthcare against unwanted access is critical. Strict policy and technological security measures must also be adopted to transmit health data to authorized people, corporations, and apps. The creation of an effective algorithm for coordination among protection, detection, and reaction services to avoid various attacks, threats, and vulnerabilities is still in progress.

A future study should investigate extracting enhanced dataset dimensions and traits, such as a new combination of vowel and gender separation. In addition, experimenting with various CNN models and training may help ameliorate methods of speech disorder detection. The allocation of samples with vocal problems is significantly uneven, complicating the diagnosis of voice diseases. A specific type of voice pathology may exist only once in the entire dataset, such as SVD. Therefore, training a specific type of vocal pathology was difficult, which led to low accuracy. Moreover, methods used in speech pathology diagnosis must be highly accurate to prevent any possible errors in the classification procedure. Optimal feature extraction and classifier must be chosen to obtain accurate and effective results of voice pathology detection.

Conflict of interest: Authors state no conflict of interest.

References

[1] AL-Dhief FT, Latiff NMAA, Malik NNNA, Sabri N, Baki MM, Albadr MAA, et al. Voice pathology detection using machine learning technique. 2020 IEEE 5th International Symposium on Telecommunication Technologies (ISTT). Manhattan, New York, USA: IEEE; 2020. p. 99–104.

[2] Mohammed MA, Abdulkareem KH, Mostafa SA, Khanapi Abd Ghani M, Maashi MS, Garcia-Zapirain B, et al. Voice pathology detection and classification using convolutional neural network model. Appl Sci. 2020;10(11):3723.

[3] Subathra MSP, Mohammed MA, Maashi MS, Garcia-Zapirain B, Sairamya NJ, George ST. Detection of focal and non-focal electroencephalogram signals using fast walsh-hadamard transform and artificial neural network. Sensors. 2020;20(17):4952.

[4] Al-Nasheri A, Muhammad G, Alsulaiman M, Ali Z, Malki KH, Mesallam TA, et al. Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions. IEEE Access. 2018;6:6961–74. doi: 10.1109/ACCESS.2017.2696056.

[5] Islam R, Tarique M, Abdel-Raheem E. A survey on signal processing based pathological voice detection techniques. IEEE Access. 2020;8:66749–76. doi: 10.1109/ACCESS.2020.2985280.

[6] Fakoor R, Ladhak F, Nazi A, Huber M. Using deep learning to enhance cancer diagnosis and classification. Proceeding of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA. Vol. 28; 2013

[7] Mansoor A, Cerrolaza JJ, Idrees R, Biggs E, Alsharid MA, Avery RA, et al. Deep learning guided partitioned shape model for anterior visual pathway segmentation. IEEE Trans Med Imaging. 2016;35(8):1856–65. doi: 10.1109/TMI.2016.2535222.

[8] Shan J, Li L. A deep learning method for microaneurysm detection in fundus images. IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); 2016. p. 357–8. doi: 10.1109/CHASE.2016.12.

[9] Fritscher K, Raudaschl P, Zaffino P, Spadea M, Sharp G, Schubert R. Deep neural networks for fast segmentation of 3D medical images. International Conference on Medical Image Computing and Computer-Assisted Intervention. Vol. 9901. 2016. p. 158–65.

[10] Cummings JL, Benson DF, Hill MA, Read S. Aphasia in dementia of the alzheimer type. Neurology. 1985;35(3):394–4. doi: 10.1212/wnl.35.3.394.

[11] Forbes K, Shanks AMV. Distinct patterns of spontaneous speech deterioration: An early predictor of Alzheimer's disease. Brain Cognition. 2002;48(2–3):356–61. doi: 10.1006/brcg.2001.1377.

[12] Johns Hopkins Medicine, 2022, Voice disorders, 10 November 2021, Voice Disorders | Johns Hopkins Medicine, Baltimore, Maryland.

[13] Muhammad G, Alhussein M. Convergence of artificial intelligence and internet of things in smart healthcare: a case study of voice pathology detection. IEEE Access. 2021;9:89198–209. doi: 10.1109/ACCE.

[14] Hegde S, Shetty S, Rai S, Dodderi T. A survey on machine learning approaches for automatic detection of voice disorders. J Voice. 2019;33:947.e11–33. doi: 10.1016/j.jvoice.2018.07.014.

[15] Al-nasheri A, Muhammad G, Alsulaiman M, Ali Z, Mesallam T, Farahat M, et al. An investigation of multi-dimensional voice program parameters in three different databases for voice pathology detection and classification. J Voice. 2017;31:113.e9–18. doi: 10.1016/j.jvoice.2016.03.019. [online] Voice and Speech Laboratory | Mass. Eye and Ear (masseyeandear.org).

[16] Kay Elemetrics Corp., Disordered Voice Database, Version 1.03 (CD-ROM), MEEI, Voice and Speech Lab, Boston, MA; October 1994.

[17] Saenz-Lechon N, Godino-Llorente JI, Osma-Ruiz V, Gomez-Vilda P. Methodological issues in the development of automatic systems for voice pathology detection. Biomedical Signal Processing and Control. 2006;1(2):120–8.

[18] Barry WJ, Pützer M. Saarbrucken voice database. Institute of Phonetics, University of Saarland. http://www.stimmdatenbank.coli.uni-saarland.de/

[19] Roy N, Merrill RM, Thibeault S, Parsa RA, Gray SD, Smith EM. Prevalence of voice disorders in teachers and the general population. J Speech Lang Hear Res. 2004;47(2):281–93.

[20] Sáenz-Lechón N, Godino-Llorente JI, Osma-Ruiz Ví, Gómez-Vilda P. Methodological issues in the development of automatic systems for voice pathology detection. Biomed Signal Process Control. 2006;1(2):120–8.

[21] Mesallam T, Farahat M, Malki K, Alsulaiman M, Ali Z, Al-nasheri A, et al. Development of the arabic voice pathology database and its evaluation by using speech features and machine learning algorithms. J Healthc Eng. 2017;2017:13. doi: 10.1155/2017/8783751. (ksu.edu.sa).

[22] Muhammad G, Alhamid M, Hossain M, Almogren A, Vasilakos A. Enhanced living by assessing voice pathology using a co-occurrence matrix. Sensors. 2017;17:267. doi: 10.3390/s17020267.

[23] Muhammad G, Alsulaiman M, Ali1 Z, Mesallam T, Farahat M, Malki K, et al. Voice pathology detection using interlaced derivative pattern on glottal source excitation. Biomed Signal Process Control. 2017;31:156–64.
[24] Al-nasheri A, Muhammad G, Alsulaiman M, Ali Z, Mesallam T, Farahat M, et al. Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions. IEEE Access. 2017;6:6961–74. doi: 10.1109/ACCESS.2017.2696056.
[25] Alhussein M, Muhammad G. Voice pathology detection using deep learning on mobile healthcare framework. IEEE Access. 2018;6:41034–41. doi: 10.1109/ACCESS.2018.2856238.
[26] Ali Z, Hossain M, Muhammad G, Sangaiah A. An intelligent healthcare system for detection and classification to discriminate vocal fold disorders. Future Gener Computer Syst. 2018;85:19–28. doi: 10.1016/j.future.2018.02.021.
[27] Rueda A, Krishnan S. Augmenting dysphonia voice using fourier-based synchrosqueezing transform for a cnn classifier. ICASSP(IEEE). 2019;6415–9.
[28] Alhussein M, Muhammad G. Automatic voice pathology monitoring using parallel deep models for smart healthcare. IEEE Access. 2019;7:46474–79. doi: 10.1109/ACCESS.2019.2905597.
[29] Hammami I, Salhi L, Labidi S. Voice pathologies classification and detection using EMD-DWT analysis based on higher order statistic features. IRBM. 2020;41:161–71. doi: 10.1016/j.irbm.2019.11.004.
[30] Fonsecaa E, Guidoa R, Junior S, Dezani H, Gati R, Pereira D. Acoustic investigation of speech pathologies based on the discriminative paraconsistent machine (DPM). Biomed Signal Process Control. 2020;55:101615. doi: 10.1016/j.bspc.2019.101615.
[31] Harar P, Galaz Z, Alonso-Hernandez J, Mekyska J, Burget R, Smekal Z. Investigation of supervised deep learning, gradient boosting, and anomaly detection approaches across four databases. Neural Comput Appl. 2020;32:15759–59. doi: 10.1007/s00521-019-044692.
[32] Syed S, Rashid M, Hussain S, Zahid H. Comparative analysis of CNN and RNN for voice pathology detection. BioMed Res Int. 2021;2021:1–8. doi: 10.1155/2021/6635964.
[33] Al-dhief F, Baki M, Latiff N, Malik N, Salim N, Albader M, et al. Voice pathology detection and classification by adopting online sequential extreme learning machine. IEEE Access. 2021;9:77293–306. doi: 10.1109/ACCESS.2021.3082565.
[34] Dave N. Feature extraction methods LPC, PLP and MFCC in speech recognition. Int J Advance Res Eng Technol. 2013;1(VI):1–5.
[35] Xie L, Liu Z. A comparative study of audio features for audio to visual cobversion in MPEG-4 COMPLIANT FACIAL ANImation. Proc. of ICMLC, Dalian; 2006. p. 13–6.
[36] Leong A. A music identification system based on audio content similarity. Thesis of Bachelor of Engineering, Division of Electrical Engineering, The School of Information Technology and Electrical Engineering, The University of Queensland; 2003.
[37] Alan V, Schafer RW. Fourier transform and Fourier analysis of signals using the discrete Fourier transform. Discrete-time signal processing. 3rd edn. London, U.K.: Pearson; 2009. p. 855–9.
[38] Everthon S, Capobianco RG, Sylvio B, Henrique D, Rodrigo R, Denis C. Acoustic investigation of speech pathologies based on the discriminative paraconsistent machine (DPM). Biomed Signal Process Control. 2020;55:101615.
[39] Cordeiro H, Ribeiro C. Spectral envelope first peak and periodic component in pathological voices. A Spectr Anal Proc Computer Sci. 2018;138:64–71.
[40] Rusz J, ˇcka J, Tykalová T, Novotný M, Dušek P, Šonka K, et al. Smartphone allows capture of speech abnormalities associated with high risk of developing parkinson's disease. IEEE Trans Neural Syst Rehab Eng. 2018;26:1495–507.
[41] Laaridh I, Meunier C, Fredouille C. Perceptual evaluation for automatic anomaly detection in disordered speech: Focus on ambiguous cases. Speech Commun Elsevier. 2018;105:23–33.
[42] Ali Z, Muhammad G, Alhamid M. An automatic health monitoring system for patients suffering from voice complications in smart cities. Access IEEE. 2017;5:3900–8.
[43] Albadr MAA, Tiun S. Spoken language identification based on particle swarm optimisation–extreme learning machine approach. Circuits Syst Signal Process. 2020;39(9):4596–622.
[44] Albadra M, Tiuna S. Extreme learning machine: A review. Int J Appl Eng Res. 2017;12(14):4610–23.
[45] Huang G, Liang N, Rong H, Saratchandran P, Sundararajan N. On-line sequential extreme learning machine. Proceedings of IASTED International Conference of Computational Intelligence; 2005. p. 232–7.
[46] Nica A, Caruntu A, Toderean G, Buza O. Analysis and synthesis of vowels using matlab. IEEE Conference on Automation, Quality and Testing, Robotics. Vol. 2. 2006. p. 371–4, 25–28.
[47] Yuhas B, Goldstein M Jr, Sejnowski T, Jenkins R. Neural network models of sensory integration for improved vowel recognition. Proc IEEE. 1990;78(10):1658–68.
[48] Buza O, Toderean G, Nica A, Caruntu A. Voice signal processing for speech synthesis. IEEE International Conference on Automation, Quality and Testing Robotics. Vol. 2. 2006. p. 360–4, 25–28.
[49] Honig F, Stemmer G, Hacker C. Brugnara, fabio, revising perceptual linear prediction. Interspeech-2005. 2005;2997–3000.
[50] Hermansky H. Perceptual linear predictive (PLP) analysis of speech. Acoustical Soc Am J. Apr. 1990;87:1738–52.

[51] Pradhan M, Minz S, Shrivastava V. Fisher discriminant ratio based multiview active learning for the classification of remote sensing images. Proceedings of the 4th IEEE International Conference on Recent Advances in Information Technology, RAIT. 1018, 2018. p. 1–6.
[52] Wang S, Li D, Wei Y, Li H. A feature selection method based on fisher's discriminant ratio for text sentiment classification. WISM. 2009;106:LNCS 5854, 88–97–501.
[53] de Sa VR. Learning classification with unlabeled data. Proc Adv Neural Inf Process Syst. 1994;6:112–9.
[54] Hossain M, Muhammad G, Alamri A. Smart healthcare monitoring: a voice pathology detection paradigm for smart cities. Multimed Syst. 2017;25:565–75. doi: 10.1007/s00530-017-0561-x.
[55] Roy S, Sayim M, Akhand M. Pathological voice classification using deep learning. CASERT. 2019;2019:1–6.
[56] Ghoniem R. Deep genetic algorithm-based voice pathology diagnostic system deep genetic algorithm-based voice pathology diagnostic system. Researchgate. 2019;11608:220–33. doi: 10.1007/978-3-030-23281-8_18.
[57] Al-Dhief F, Latiff N, Malik N, Salim N, Baki M, Albadr M, et al. A survey of voice pathology surveillance systems based on internet of things and machine learning algorithms. IEEE Access. 2020;8:64514–33. doi: 10.1109/ACCESS.2020.2984925.
[58] Narendra N, Alku P. Glottal source information for pathological voice detection. IEEE Access. 2020;8:67745–55.
[59] Tuncer T, Dogan S, Özyurt F. Novel multi center and threshold ternary pattern based method for disease detection method using voice. IEEE Access. 2020;8:84532–40.
[60] Lee J. Experimental evaluation of deep learning methods for an intelligent pathological voice detection system using the saarbruecken voice database. Appl Sci. 2021;11:7149.
[61] Fan Z, Wu Y, Zhou C, Zhang X, Tao Z. Class-imbalanced voice pathology detection and classification using fuzzy cluster oversampling method. Appl Sci. 2021;11:3450.
[62] Marjani M, Nasaruddin F, Gani A, Karim A, Hashem I, Siddiqa A, et al. Big IoT data analytics: architecture, opportunities, and open research challenges. IEEE Access. 2017;5:5247–61.
[63] Al Nuaimi E, Al Neyadi H, Mohamed N, Al-Jaroodi J. Applications of big data to smart cities. J Internet Serv Appl. 2015;6:25.
[64] Gubbi J, Buyya R, Marusic S, Palaniswami M. Internet of Things (IoT): A vision, architectural elements, and future directions. Future Gener Comput Syst. 2013;29(7):1645–60.
[65] Atlam H, Walters R, Wills G. Fog computing and the Internet of Things: A review. Big Data Cogn Comput. 2018;2(2):10.
[66] Li X, Wang Q, Lan X, Chen X, Zhang N, Chen D. Enhancing cloud-based IoT security through trustworthy cloud service: An integration of security and reputation approach. IEEE Access. 2019;7:9368–83.
[67] Botta A, de Donato W, Persico V, Pescapé A. Integration of cloud computing and Internet of Things: A survey. Future Gener Comput Syst. 2016;56:684–700.
[68] Bhattacharyya N. The prevalence of voice problems among adults in the united states. Laryngoscope. 2014;124(10):2359–62.
[69] Muhammad G, Alhamid MF, Alsulaiman M, Gupta B. Edge computing with cloud for voice disorder assessment and treatment. IEEE Commun Mag. 2018;56(4):60–5.
[70] Mohammadi M, Al-Fuqaha A. Deep Learning for IoT Big Data and Streaming Analytics: A Survey. IEEE Commun Surv Tutor. 2018;20:2923–60. doi: 10.1109/COMST.2018.2844341.