A Survey on Machine Learning Approaches for Automatic Detection of Voice Disorders

[*]Sarika Hegde , [*]Surendra Shetty, [*]Smitha Rai, and [†]Thejaswi Dodderi, *Udupi, and †Mangaluru, India

Summary: The human voice production system is an intricate biological device capable of modulating pitch and loudness. Inherent internal and/or external factors often damage the vocal folds and result in some change of voice. The consequences are reflected in body functioning and emotional standing. Hence, it is paramount to identify voice changes at an early stage and provide the patient with an opportunity to overcome any ramification and enhance their quality of life. In this line of work, automatic detection of voice disorders using machine learning techniques plays a key role, as it is proven to help ease the process of understanding the voice disorder. In recent years, many researchers have investigated techniques for an automated system that helps clinicians with early diagnosis of voice disorders. In this paper, we present a survey of research work conducted on automatic detection of voice disorders and explore how it is able to identify the different types of voice disorders. We also analyze different databases, feature extraction techniques, and machine learning approaches used in these research works.

Key Words: Voice disorder−Pathological voice−Machine learning−Automatic detection−Literature survey −Review.

INTRODUCTION

Speech is the basic human instinct and voice one among its subsystem. Recently, the study of vocalization is coined as "Vocology."[1] Normal voice is the acoustic product of pulmonary air pulses interacting with the larynx that sets the adduction of true vocal folds and produces periodic and/or aperiodic sounds. Often, multiple abusive vocal habits, commonly referred as vocal hyperfunction, result in voice disorders like aphonia (complete loss of voice) and/or dysphonia (partial loss of voice). To define, voice disorder is anything that deviates "quality, pitch, loudness, and/or vocal flexibility" from voices of similar age, gender, and social groups.[2] Effect of nonmalignant voice disorders are not life threatening, but impact of an untreated voice disorder can significantly impact social, professional, and personal aspects of communication.[3] Clinically, the foundation of voice therapy is to optimize laryngeal and supralaryngeal muscle tension during voice production in an attempt to improve vocal efficiency and also enhance resolution of the vocal fold lesion. Changes in lifestyle and professional demands have increased public awareness of the need for voice care. Citizens now make frequent visits to laryngologists and voice therapists for maintaining vocal health. During such visits, it is common to undergo invasive tests like video or laryngoscopy/stroboscopy, which are both difficult to administer and expensive.

Hence, profiling acoustical−perceptual features would serve as the efficient tool to compare performance variations before and after voice therapy and also adjunct development of voice database for automated voice recognition systems. These issues instigated more research in voice recording and voice analysis. The outcome of this is automatic detection. An automatic voice disorder detection (AVDD) system can be helpful for both patients and laryngologists. The patient can use the system for early detection of voice disorder and the voice pathologist for rehabilitation decisions.

The objective of this article is to discuss and present a detailed analysis of the various voice disorders and machine learning (ML) techniques used so far, for developing automatic systems. The article is organized as follows: in "Voice Disorders" section, we discuss various types of voice disorders. "AVDD System" section describes the architecture of typical AVDD system; "Voice Disorder Database" section describes some voice disorder databases that are commonly used for this kind of research; then, we explore different feature extraction techniques in "Feature Extraction Techniques" section; various ML approaches used for voice disorder detection are discussed in "ML Algorithms" section; we summarize the research works and present a conclusion in "Discussions" and "Conclusions" sections, respectively.

VOICE DISORDERS

Communication science disorders are classified into five categories: speech disorders, language disorders, social communication disorders, cognitive communication disorders, and swallowing disorders.[4] As per the Classification Manual for Voice Disorders, hyper- and hypofunctional vocal fold pathologies are categorized as follows: (i) structural lesions, like vocal nodules, vocal polyps; (ii) inflammatory conditions like acute laryngitis; (iii) trauma or injury based; (iv) systemic conditions, for example, hyperthyroidism and hypothyroidism; (v) nonlaryngeal aerodigestive disorders, consisting of reflux disorder and bronchitis among others;

Accepted for publication July 10, 2018.
From the *NMAM Institute of Technology, Udupi, Karnataka, India; and the †Nitte Institute of Speech & Hearing, Mangaluru, Karnataka, India.
Address correspondence and reprint requests to Sarika Hegde, CSE Department, NMAM Institute of Technology, Nitte, Karkala, Udupi, Karnataka, India. E-mail: sarika.hegde@yahoo.in
Journal of Voice, Vol. 33, No. 6, pp. 947.e11−947.e33
0892-1997

https://doi.org/10.1016/j.jvoice.2018.07.014

(vi) psychiatric and psychological disorders, which are functional voice disorder and/or mutism; (vii) neurological conditions, for example, adductor palsy, abductor palsy, and spasmodic dysphonia; (viii) other disorders like muscle tension dysphonia; and (ix) undiagnosed and otherwise not specified.[5]

Among different vocal fold lesions, mass-based pathologies are highly prevalent due to the phonotraumatic impact it delivers on the susceptible multilayered vocal folds. Persistent tissue inflammation and external influences often results in vocal nodule and vocal polyp.[6] Under these circumstances the vocal fold closure is incomplete, the voice production is not economical and perceptually hoarse. On the contrary, in nonphonotraumatic voice disorders, like muscle tension dysphonia and functional voice disorder, there is no vocal fold lesions, but vocal fatigue, degraded voice quality, and increased laryngeal tension are observable.

Multiparametric assessment protocol is considered ideal for voice evaluation.[2,7] Traditionally, a structured approach is essential and it encompasses the following: patient interview, laryngeal imaging by stroboscopy and/or laryngoscopy, basic aerodynamic assessment, perceptual analysis by standardized psychoacoustic methods, acoustical analysis, and subjective voice evaluation. Considering the recent technological advancement, voice scientists have been forefront in developing tools for acoustical analysis to differentiate normal voice from those with aphonia and/or dysphonia.

The aim of such innovations must be to offer economical solutions that is sensitive to identify voice changes and remains reliable across dynamic clinical setups. With due research in acoustical recording and analysis, the outcome is automatic detection.

AVDD SYSTEM

Systems for the automatic detection of voice disorders can be designed and developed using ML algorithms. Here, the voice data need to be preprocessed and converted into a set of features before applying an ML algorithm. The important steps to be followed are shown in Figure 1:

1. A given set of voice data, in audio files, need to be labeled manually by experts as representing healthy or a pathological voice.
2. Then, the raw audio data in each file are divided into short frames and each frame is processed to extract features from it.
3. The set of features extracted from all the frames are considered as input for the ML algorithms.

The dataset is divided into training and testing sets by randomly selecting the observations belonging to both normal and pathological voices. The training set is used to develop the ML model and the testing set is used to evaluate the model. During the evaluation process, the classification accuracy is computed. This classification accuracy is taken as a metric for assessing the performance of various AVDD systems.

VOICE DISORDER DATABASE

Voice disorder database is an essential element in AVDD system. The dataset consists of voice recordings of both normal and pathological voices. The recordings can contain either sustained vowel phonation or continuous speech. In our survey, we have observed that most of the researchers have used standard databases, such as Massachusetts Eye and Ear Infirmary (MEEI), Saarbruecken Voice Database (SVD), and Arabic Voice Pathology Database (AVPD). The more details about these three databases can be found in Al-nasheri et al.[8] Some of the researchers have used private dataset developed in collaboration with local hospitals.

FEATURE EXTRACTION TECHNIQUES

Feature extraction is the first step in any voice disorder detection system. In this step, the given voice signals are converted into representative acoustic features using various digital signal processing techniques. We will now discuss the most popularly used techniques for acoustic analysis and feature extraction in the related area.

Acoustic analysis

Acoustic analysis is the measurement of sound information in voice. The results of acoustic analysis could be used for measuring the severity of a voice disorder. Some of the measures associated with acoustic analysis of voice signals are as follows:

(i) Perturbations in the fundamental pitch period and peak amplitude.
(ii) Vocal noise included in the signal.
(iii) Cycle-to-cycle wave form variations.
(iv) Average frequency characteristics.
(v) Transition characteristics of the signal.

The application of acoustic analysis in voice disorder detection can be found in Kasuya et al[9] and Sonu and Sharma.[10] Multidimensional Voice Program (MDVP) is standard software for acoustic analysis and it is most popularly used in AVDD systems. It is reliable and very comprehensive. Using MDVP, one can estimate 33 voice parameters, which measure frequency related, intensity related, noise related, tremor related, and its perturbations measures among others.[11]

Mel-frequency cepstral coefficient

Mel-frequency cepstral coefficient (MFCC) is a standard method for feature extraction that makes use of the knowledge of human auditory system. The general steps for extracting the MFCC features for a single frame[12,13] are as follows:

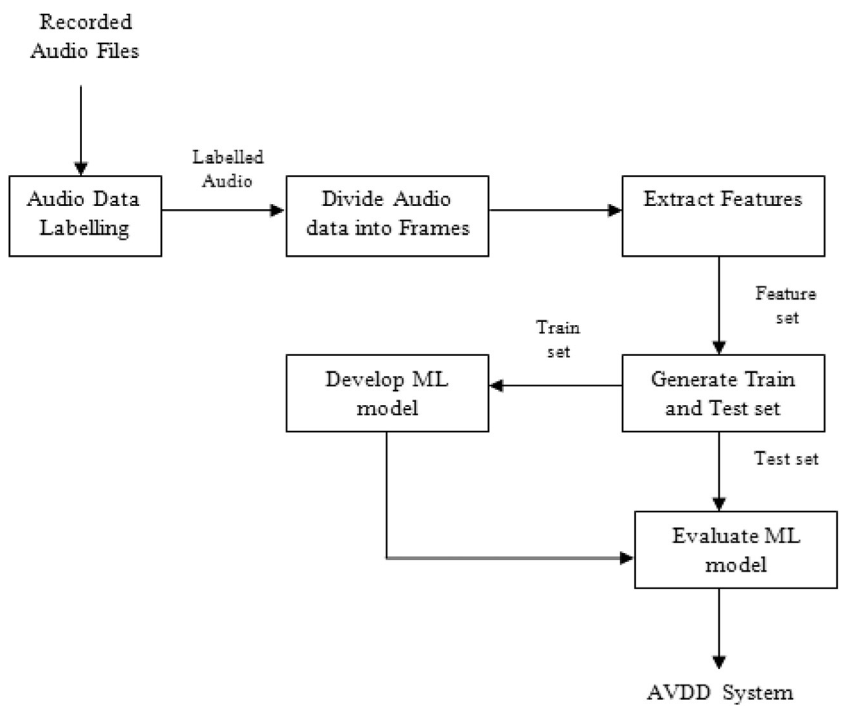

FIGURE 1. Automatic voice disorder detection (AVDD) system.

i. Computation of the discrete Fourier transform coefficients.
ii. Filtering with Mel spaced triangular filter.
iii. Computation of sub-band energies.
iv. Computation of the discrete cosine transform coefficients.

Linear predictive coefficients

In linear predictive (LP) analysis, the first step is to estimate the source signal using inverse filtering. After that, the source signal is used to compute the spectrum. The computed spectrum is used to study the energy distribution in both normal and pathological voices. The number of LP coefficients is one of the key factors of the LP analysis in order to determine the formant peaks. Because, removing the effect of the formants from the speech signal will provide an accurate estimation of the source signal.[14]

Discrete wavelet transform

The discrete wavelet transform (DWT) can be used in variety of signal processing applications. It is discrete in time and scale. The voice is a time variant signal and sometimes need to be converted into frequency domain for the analysis. DWT is capable of performing a joint time−frequency analysis and also the analysis of high frequency characteristics of the pathological voices. Hence, it is useful tool for voice disorder detection. The DWT coefficients may have real values, but the time and scale values used to index these coefficients are integers.[82,93]

Glottal flow signal parameters

The glottal flow signal can be obtained by performing an inverse filtering of the voice signal, which consists of eliminating the influence of the vocal tract and the voice radiation caused by the mouth, and preserving the glottal flow signal characteristics.[15] One of the toolboxes in MATLAB, called Aparat8, was developed especially based on the Pitch Synchronous Iterative Adaptive Inverse Filtering algorithm to obtain the glottal signal parameters. The time-domain parameters that can be extracted from glottal flow signal are: (i) closing phase, (ii) opening phase, (iii) open quotient, (iv) closed quotient, (v) amplitude quotient, (vi) normalized amplitude quotient, (viii) quasi-open quotient, and (ix) speed quotient. The frequency domain parameters are: (i) difference between harmonics and (ii) harmonics richness factor.

Feature selection/reduction/transformation algorithms

Feature reduction is a preprocessing step in ML. The main reason for this is to eliminate repeated and inappropriate features in order to improve the effectiveness of the system. The selection of highly useful features from a dataset will reduce the complexity of the learning phase and thereby improves the ability of the classifiers to generalize.[16] Some of the most commonly used dimensionality reduction techniques are as follows:

(i) Principal component analysis (PCA)
(ii) Linear discriminant analysis (LDA)
(iii) Genetic algorithm
(iv) Higher order singular value decomposition
(v) Relief
(vi) Minimum redundancy maximum relevance
(vii) Fisher discrimination ratio (FDR)

ML ALGORITHMS

Research works in the area of pathological voice detection and classification started in early 1980s when pattern recognition and ML techniques were not much advanced and popular. There are few works carried out using basic techniques like distance measure, vector quantization (VQ), and statistical analysis, etc. However, in recent days, ML algorithms are popularly used for detection of pathological voice based on the computed acoustic features of the input signal. In this section, we will first discuss the significant research works that has not used specific ML algorithms. Later, we discuss the research works based on ML techniques categorized into hidden Markov model (HMM), Gaussian mixture model (GMM), support vector machine (SVM), artificial neural networks (ANN), decision trees, linear classifier, K-means clustering, and combined classifiers.

Deller and Anderson[17] have used inverse filtering for classification and assessment of laryngeal dysfunction using the roots of digital inverse filter. The z-plane roots and a vector of pattern features are further processed by an automatic clustering procedure. They were able to demonstrate the ability of inverse filter technique to identify the simulated anomalous laryngeal behavior reflected in the speech signal. Childers and Bae[18] have developed two methods for the detection of laryngeal pathology: (1) a spectral distortion measure, using pitch synchronous and asynchronous methods with linear predictive coding (LPC) vectors and VQ; and (2) analysis of the Electroglottograph (EGG) signal using time interval and amplitude difference measures. These two methods achieved 75.9% and 69.0% probability of pathological detection respectively. Cairns et al[19] have proposed a noninvasive method to detect the hypernasality in speech based on a nonlinear operator called Teager Energy Operator. They classified the normal and hypernasal voices based on probability distribution function. The maximum classification accuracy was 94.7%. Accardo and Mumolo[20] described an algorithm using fractal dimension parameters, energy ratio and zero-crossing features. Distance matrix of those features is used for comparing the normal and pathological voices. The fractal dimension, energy ratio, and zero crossing features obtained classification accuracy of 96.1%, 92.1%, and 94.1%, respectively.

Parsa and Jamieson[21] investigated glottal noise measures (signal-to-noise ratio, harmonic-to-noise ratio [HNR], normalized noise energy, frequency domain HNR, pitch amplitude, and spectral flatness ratio) as features for classifying healthy and pathological voices. They classified the measures of two different classes by comparing (i) probability distribution of measures, (ii) ranking of measures, (iii) Receiver Operating Characteristic of each measure. The highest classification rate obtained was 96.5%. Hadjitodorov et al[22] proposed an approach based on modeling of the probability density functions of the input vectors of the normal and pathological speakers by means of prototype distribution maps (PDM) and achieved 95.1% classification accuracy. The features like, pitch period, shape of the pitch pulse, HNR, low-to-high energy ratio, etc. are used. de Oliveira Rosa et al[23] used the residue (extracted using inverse filtering) and some acoustic features with a statistical method to discriminate healthy from pathological voices. The best discrimination ability obtained was 54.79% using jitter and prototype distribution neural map—PDM(X). Watts et al[24] evaluated the fundamental frequency, jitter, shimmer, long-term frequency, and amplitude variability of a voice of a professional singer with vocal fold edema before and after medication. They observed an increase in fundamental frequency (f_0) and large decreases in jitter, shimmer, long-term frequency, and amplitude variability after medication. Guido et al[25] compared the performances of different DWTs like Haar, Daubechies, Coiflets, Symmlets, and Spikelet in discriminating normal and pathological voices using a threshold method. The highest classification accuracy obtained was 90% with Spikelet. Umapathy et al[26] proposed a joint time−frequency analysis approach for classifying pathological voices, in which speech signals were decomposed using an adaptive time−frequency transform algorithm, and several features such as the octave max, octave mean, energy ratio, length ratio, and frequency ratio were extracted from the decomposition parameters and analyzed using statistical pattern classification techniques. The classification accuracy achieved was 93.4%. Zhang et al[27] have concluded in their paper that combining traditional perturbation analysis and nonlinear dynamic analysis might provide efficient descriptions of pathological voices and represent a valuable tool for clinical diagnosis of laryngeal paralysis.

Neto et al[28] developed a noninvasive method for detecting vocal fold edema using LPC, LPC-based cepstral coefficients, and MFCC with a vector-quantizing-trained distance classifier. They achieved classification accuracies of up to 85%, 80%, and 52% respectively with these features. Gómez-Vilda et al[29,30] described a method for voice pathology detection using biomechanical parameters with statistical methods like hierarchical clustering, PCA, and LDA. They claim that biomechanical parameters combined with acoustic parameters can be efficiently used to detect voice pathology. Zhang and Jiang[31] investigated the acoustic characteristics of sustained and running vowels from normal subjects and patients with laryngeal pathologies. Perturbation methods (including jitter and shimmer), signal-to-noise ratio, and nonlinear dynamic methods (such as correlation dimension and second-order entropy) were used to analyze sustained and running vowels. Mann-Whitney rank sum tests showed a statistically significant difference between pathological and normal voices. Fontes et al[32] have illustrated that the existing feature extraction and classification for voice pathology detection are complex and hence, not feasible to be applied for real-time system. They proposed a novel simple low-complexity approach for the detection and classification of laryngeal pathologies using a feature, correntropy spectral density. The voice samples from MEEI database are classified using a simple measure of Euclidian distance with a success rate of 97%.

In the next section, we discuss the research works carried based on the specific ML algorithms. All the papers are summarized in Table 1 with the details as author names, database used, list of features extracted, type of voice disorder detected, ML algorithm used, and the accuracy. We discuss only the significant research works based on accuracy (above 90%) and highlight the same in the table.

HMM

HMM is used to model the spectral variability of each speech sound using a mixture Gaussian distribution.[33–35] HMM acts as a stochastic finite state machine, which is assumed to be built from a finite set of possible states (hidden during evaluation) and each of those states are associated with a mixture of Gaussian probability density function.[36]

The study in Gavidia-Ceballos and Hansen[37] has proposed a method for vocal fold cancer detection using HMM. They have discussed the computation of new features characterizing pathology, termed enhanced spectral-pathology component, which does not require the estimation of glottal flow waveform as in other works in the literature. An iterative procedure-based estimation−maximization is used to fit the features into a stochastic model, HMM is evaluated on the speech recordings from healthy and vocal fold cancer patients of sustained vowel sounds. The highest accuracy for healthy and pathological voices is 92.8% and 88.7%, respectively, with finer spectral representation of enhanced spectral-pathology component. Similarly, HMM is used by Arias-Londoño et al[38] for feature space transformation of MFCC and short-term noise parameter features. The main drawbacks of conventional feature space transformation (such as PCA and MDA) are inconsistencies in feature extraction and classification stages, and these methods do not take into account the temporal dependencies among the observations, considering the observations independent. To overcome this, new technique is proposed in which transformation and classification stage can be obtained simultaneously and the parameters of the model can be adjusted with a criterion that minimizes the classification error. The technique is applied on MEEI database with an accuracy of 96.61%.

GMM

GMM is used to model the probability distribution of feature values, which are continuous in nature. The voice acoustic measures are continuous. Hence, it is suitable to model the voice characteristics. GMM parameters are estimated during training using expectation−maximization iterative algorithm or maximum *a posteriori* algorithm.

Muhammad et al[39] have used GMM in their paper for classification of voice recordings (Arabic Digits) of patients with six types of voice disorders, namely vocal fold cysts, laryngopharyngeal reflux disease, spasmodic dysphonia, sulcus vocalis, vocal fold nodules, and vocal fold polyps. They illustrated that the important features like voice onset and voice offset attributes, which contributes significantly in detection of voice disorders, are ignored in case of sustained vowels input. Hence, they worked on continuous speech domain rather than sustained vowels and have proposed a novel feature extraction method, multidirectional regression, which takes into account consonant and vowel locations, formant transitions, and voice onset and offset distributions for continuous speech. The features are classified using GMM with 99% accuracy. Similarly, Ali et al[40] have proposed a method for pathological voice detection and pathology classification using GMM classifier for running speech voice dataset from MEEI database. They have computed auditory spectrum and all-pole model-based cepstral coefficients features based on the set of three psychophysics conditions of hearing (critical band spectral estimation, equal loudness haring curve, and the intensity loudness power of law of hearing). The features are visualized and analyzed using GMM classifier to differentiate between normal and pathological voices and also further classified into types like adductor, keratosis, vocal nodules, vocal fold polyp, and paralysis. The highest accuracy for pathology detection and pathology classification with auditory spectrum is 93.33% (adductor), respectively. Accuracy for the same with all-pole model-based cepstral coefficients is 99.56% and 89.47%, respectively. The important observations made in their work are as follows. Even though the literature works have demonstrated a very good accuracy for database with sustained vowels, the amount of work carried on running speech is less and also it is a challenging task. They have stressed on the importance of research work with continuous speech database as it is more realistic due to the usage in conversations in daily life. The processing of continuous speech requires voice activity detection (VAD) activity, which is most challenging task and hence, results in low performance. The authors have proposed features that can be computed without VAD. Based on the literature survey, they have also stated that MFCC is a good feature for pathology detection but not in the case of pathology classification. Overall, they have claimed to propose a method that outperforms the performance of existing running speech database experiments without the need of VAD. They also claim that the features are visually appealing and could be used by clinicians directly for assessment, making the use of GMM classifier as optional.

Ali et al[41] have proposed a system for voice disorder detection using GMM, where they detect the voice disorder by determining the source signal from the speech through the linear prediction analysis. The spectrum computed using the features obtained from LP analysis provides the distribution of energy in both normal and pathological voices, which will be used to differentiate them. They have observed that lower frequencies from 1 to 1562 Hz contribute significantly in the detection of voice disorders. The system is tested with both sustained vowels and running speech and achieved accuracy of 99.94% and 99.75 ± 0.8, respectively.

TABLE 1.
Summary of Research Works Related to Voice Disorder Detection and Classification

Sl. No.	Author	Database	Feature Extraction Technique	Voice Disorder Type	ML Technique	Best Accuracy (%)
1	Gavidia-Ceballos and Hansen[37]	Sustained vowel phonation	Enhanced spectral-pathology component	Vocal fold cancer	HMM	92.8 for healthy and 88.7 for pathological
2	Martinez and Rufiner[83]	Not mentioned	Cepstrum, mel-cepstrum, delta cepstrum and delta mel-cepstrum, and Fast Fourier Transform (FFT)	Not mentioned	ANN	91.30
3	Dibazar et al[84]	MEEI	MFCC and measures of pitch dynamics	Adductor spasmodic dysphonia, A–P squeezing, paresis, gastric reflux, hyper-function paralysis, keratosis/ leukoplakia	HMM	99.44%
4	Hadjitodorov and Mitev[69]	MEEI	Shimmer, jitter, and several harmonics-to-noise ratios	Not mentioned	K Nearest Neighbor (KNN)	96.1
5	Ritchings et al[60]	Christie and Withington hospitals in Manchester	Short-term + long-term parameters	Laryngeal cancer	MLP	92
6	Nayak and Bhat[85]	Kay Elemetrics Corporation	Normalized energy	Paralysis	Multilayer ANN	90
7	Ananthakrishna et al[86]	MEEI	Energy spectrum	Adductor, cyst, leukoplakia, vocal fold polyp, polyp degenerative, vocal fold edema, vocal nodule	KNN	89.19
8	Godino-Llorente and Gomez-Vilda[61]	MEEI	MFCC	Glottis cancer	LVQ MLP	96 with LVQ
9	Behroozmand and Almasganj[87]	Local database	Wavelet packet sub-band and MFCC	Edema, nodules, and polyp	ANN SVM	94.12 with Entropy
10	Fonseca et al[88]	A sustained Brazilian Portuguese phoneme /a/	Daubechies discrete wavelet transform (DWT-db)	Vocal cord nodules	SVM	95%
11	Godino-Llorente et al[89]	MEEI	Short-term cepstral and noise parameters	Not mentioned	SVM	95.0 ± 1.8
12	Moran et al[90]	MEEI	Wavelet transformation patterns	Hyperfunctioning of vocal fold	ANN	85
13	Moran et al[91]	Voice recordings from the Disordered Voice Database Model 4337 system	Pitch perturbation, amplitude perturbation, and harmonic-to-noise ratio	Neuromuscular disorders Physical abnormalities Mixed pathology voice	Linear classifier	87, 78, 61 for three types of disorders mentioned

(*Continued*)

TABLE 1. (*Continued*)

Sl. No.	Author	Database	Feature Extraction Technique	Voice Disorder Type	ML Technique	Best Accuracy (%)
14	o Schlotthauer et al[92]	Not mentioned	Acoustical measurements	Spasmodic dysphonia and muscle tension dysphonia	ANN	98.94
15	Behroozmand and Almasganj[43]	MEEI	Energy and entropy features	Unilateral vocal fold paralysis	SVM with linear kernel	100
16	Crovato and Schuck[62]	PUCRS Hospital	Wavelet packet transform	Reinke's edema, degenerative edema	ANN	87.5–100
17	Fonseca et al[93]	Voice signals with sustained Brazilian Portuguese phonemes /a/	Daubechies' discrete wavelet transforms (DWT-db), linear prediction coefficients (LPC)	Not mentioned	Least squares SVM	90
18	Kukharchik et al[94]	Local database	Modified wavelet transform	-	SVM	99%
19	Shama and Cholayya[70]	MEEI	Harmonics-to-noise ratio (HNR) measure and the critical-band energy spectrum	Adductor, paralysis , cyst leukoplakia, vocal fold polyp, polyp degenerative, vocal fold edema, vocal nodule	KNN	94.38 with energy spectrum
20	Aguiar Neto et al[95]	MEEI	LPC, Cepastal (CEP), Mel cepstral (MEL) features	Vocal fold edema, cysts, nodules, and paralysis	Vector-quantizing-trained distance classifier	98
21	Gelzinis et al[75]	University Hospital of Kaunas University of Medicine, Lithuania	Pitch and amplitude perturbation measures, cepstral energy features, autocorrelation features, and linear prediction cosine transform coefficients	Diffuse and nodular	KNN SVM	95.5 with SVM
22	Linder et al[96]	Charité Berlin, Campus Benjamin Franklin	Jitter, shimmer, standard deviation of fundamental frequency, and the glottal-to-noise excitation ratio	Not mentioned	ANN	80
23	Murugesapandian et al[97]	MEEI	Mel-scaled wavelet packet transform	Vocal fold edema, gastric reflux, and vocal fold paralysis	ANN	95.75%
24	Salhi et al[98]	National Hospital "Rabta-Tunis"	Wavelet transform	Neural and vocal pathologies (Parkinson, Alzheimer, laryngeal, dyslexia)	Neural networks	Between 70% and 100%.
25	Ghoraani and Krishnan[73]	MEEI	Time–frequency distribution	Variety of diseases	K means	98.6

(*Continued*)

TABLE 1. (*Continued*)

Sl. No.	Author	Database	Feature Extraction Technique	Voice Disorder Type	ML Technique	Best Accuracy (%)
26	Hariharan et al[99]	MEEI	Mel Frequency Band Energy Coefficients (MFBECs) combined with singular value decomposition	Variety of diseases	LDA and KNN	98.48 with LDA 99.59 with KNN
27	Kotropoulos and Arce[100]	MEEI	LPC	Vocal fold paralysis in men and edema in women	Linear classifier with reject option	68.8
28	Markaki and Stylianou[101]	MEEI and Prıncipe de Asturias (PdA)	Normalized modulation spectral features	-	SVM	94.1 with MEEI
29	Markaki and Stylianou[102]	MEEI	Modulation spectral features	Polyp, adductor, Keratosis, nodules	SVM	94.1 for pathology detection and 88.6 for pathology classification
30	Arias-Londoño et al[38]	MEEI Universidad Politécnica de Madrid (UPM)	MFCC and short-term noise parameters	Organic, neurological, and traumatic disorders. UPM contained nodules, polyps, edemas, carcinomas	HMM	96.6
31	Das[76]	National Centre for Voice and Speech, Denver, Colorado	Variable selection using regression	Parkinson's disease	Neural networks, DMneural, regression, and decision tree	92.9% with neural network
32	Markaki et al[103]	MEEI and Prıncipe de Asturias (PdA)	Combination of modulation spectral features and MFCC	Dysphonia, nodules, polyps, and edema	SVM	96.37% with MEEI
33	Arias-Londono et al[77]	MEEI	Nonlinear features	Variety of diseases	GMM + SVM	98.23
34	Arjmandi et al[78]	MEEI	Long-time features	Not mentioned	SVM	94.26
35	Carvalho et al[104]	São Carlos of the University of São Paulo, Brazil	Wavelet transform	Vocal fold nodules, Reinke's edema, and neurological dysphonia	ANN	90.53
36	de Bruijn et al[105]	Local dataset	Features of nasalance	Oral cancer	ANN	-
37	Hariharan et al[63]	MEEI	Time-domain variation	Vocal fold paralysis and edema	Probabilistic neural network	90
38	Lee et al[106]	Japanese Society of Logopedics and Phoniatrics	Higher order statistics analysis	Various types of voice disorders that lead to roughness and breathiness	Classification and regression tree (CART)	88.7
39	Markaki and Stylianou[44]	MEEI	Modulation spectrum	Vocal nodules, vocal polyp, keratosis, adductor, paralysis	SVM	94.1

(*Continued*)

TABLE 1. (*Continued*)

Sl. No.	Author	Database	Feature Extraction Technique	Voice Disorder Type	ML Technique	Best Accuracy (%)
40	Saeedi et al[45]	MEEI	Wavelet filter banks	Spasmodic dysphonia, gastric reflux, keratosis, paralysis, nodules, polyp, and edema	SVM	100
41	Uloza et al[47]	Kaunas University of Medicine, Kaunas, Lithuania	Pitch and amplitude perturbation measures, frequency features, mel-frequency features, cepstral energy features, MFCC, HNR spectral, HNR cepstral, linear prediction coefficients, and linear prediction cosine transform coefficients	Nodular and diffuse lesion	SVM	90
42	Arjmandi and Pooyan[46]	MEEI	Short-time Fourier Transform, continuous wavelet transform, and wavelet packet transform	Vocal fold cyst , vocal fold nodules, vocal fold edema, and vocal fold paralysis	SVM	100% with WPT features
43	Muhammad et al[39]	Arabic digits	Multidirectional Regression (MDR)-based features	Vocal fold cysts, laryngopharyngeal reflux disease, spasmodic dysphonia, sulcus vocalis, vocal fold nodules, and vocal fold polyps.	GMM	99
44	Tsanas et al[107]	National Center for Voice and Speech (NCVS) database	Dysphonia features	Parkinson’s disease	SVM	99
45	Ali et al[108]	King Abdul Aziz University Hospital, Riyadh, Saudi Arabia	MFCC	Cyst, polyp, sulcus vocalis, nodules, and paralysis	GMM	91.66
46	Alsulaiman et al[109]	King Saud University, Riyadh	Relative Spectral transform perceptual linear predictive	Cyst, GERD, Polyp, and sulcus	Multiclass SVM	100

(*Continued*)

TABLE 1. (*Continued*)

Sl. No.	Author	Database	Feature Extraction Technique	Voice Disorder Type	ML Technique	Best Accuracy (%)
47	Belalcazar-Bolanos et al[110]	Universidad de Antioquia in Medellín, Colombia	Harmonics-to-noise ratio (HNR), normalized noise energy (NNE), cepstral HNR (CHNR), and glottal-to-noise excitation ratio (GNE)	Parkinson's disease	KNN	66.57
48	Holi[111]	J.S.S. Hospital, Mysore	MFCC	Parkinson's disease, cerebellar demyelination, stroke and senile disease	Multilayer neural network	92
49	Kaleem et al[72]	MEEI (continuous speech)	Empirical mode decomposition	Not mentioned	Linear classifier	95.7
50	Majidnezhad and Kheidorov[112]	Belarusian Republican Centre of Speech, Voice and Hearing Pathologies	Wavelet packet decomposition and MFCC	Various diseases	ANN	91.54
51	Saldanha et al[113]	Kay Elemetrics Corporation	MFCC	Not mentioned	LDA	93.14
52	Vikram and Umarani[114]	Local database (AIISH Mysore)	Wavelet-based MFCC	Parkinson's disease, vocal fold paralysis, nodules, edema	Hybrid classifier (GMM-universal background model and SVM)	96.61
53	Akbari and Arjmandi[64]	MEEI	Wavelet-packet-based energy and entropy	A–P squeezing, gastric reflux , and hyperfunction	Multilayer neural network	96.67 with energy features 97.33% with entropy features
54	Al-nasheri et al[11]	MEEI SVD	Peak and lag features	Not mentioned	GMM	97 with MEEI
55	Cordeiro et al[115]	MEEI	LPC of 30th order	Nodules, edema, paralysis, polyps, keratosis	Decision tree	94.2
56	Fezari et al[116]	German database developed by Putzer	MFCC with energy features	Spasmodic dysphonia	GMM	79.92 for pathological and 81.90 for normal
57	El Emary et al[117]	St. Theresa's clinic of Caritas in Saarbrücken, Saarland University, Institute of acoustics in Germany.	MFCC with energy features.	Neuroloogical disorders	GMM	82.14% for pathological and 82.6 for normal
58	Hariharan et al[79]	UCI machine learning database	-	Parkinson's disease	Probablistic neural network, general regression neural network, and SVM	100

(*Continued*)

TABLE 1. (*Continued*)

Sl. No.	Author	Database	Feature Extraction Technique	Voice Disorder Type	ML Technique	Best Accuracy (%)
59	Jothilakshmi[80]	Rajah Muthiah Medical College Hospital, Annamalai University	MFCC and LPC	Cerebral palsy, dysarthria, hearing impairments, laryngectom, mental retardation, left-side paralysis, quadriparesis, stammering, stroke, tumour in vocal tract	GMM	95.74
60	Muhammad and Melhem[48]	MEEI	MPEG-7	Variety of voice disorders	SVM	100
61	Novotný et al[118]	Local database	Features describing six different articulatory aspects of speech	Parkinson's disease	SVM.	80
62	Ali et al[40]	MEEI	APS and APCC	Variety of voice disorders	GMM	99.56
63	Al-nasheri et al[11]	AVPD	MDVP parameters	Cyst, nodules, paralysis, polyp, and sulcus	SVM	81.33
64	Cordeiro et al[119]	MEEI	Mel-line spectrum frequencies	Edema, Nodules, Paralysis	GMM	77.9
65	López-de-Ipina et al[120]		Linear parameters	Alzheimer's disease	KNN MLP	87.30 90.9
66	Orozco- Arroyave et al[50]	Local	Stability measures spectral–cepstral features	Parkinson's disease, laryngeal pathologies, cleft lip and plate	Gaussian Kernel SVM	81–99 95–99
67	Rani and Holi[121]	J.S.S. Hospital, Mysore	MFCC	Parkinson's disease, cerebellar demyelination, and stroke	GMM	89.5
68	Saidi and Almasganj[49]	MEEI	M-band wavelet	Vocal fold paralysis, vocal fold paresis, nodules, polyps, and edema	SVM with a linear kernel	100
69	Salehi[122]	MEEI	Parameters of lifting scheme	Paralysis, Paresis, edema, nodules, polyp, and keratosis	SVM	98.30
70	Vásquez-Correa et al[123]	Universidad de Antioquia, in Medellín, Colombia	Voiced and unvoiced features	Parkinson's disease	SVM	64–86 (voiced) 78–99 (unvoiced)
71	Agarwal et al[124]	UCI machine learning repository	Six types of dysphonia parameters including frequency (jitter), pulse, amplitude (shimmer), voicing, harmonicity, and pitch parameters	Parkinson's disease	ELM	90.76
72	Aicha and Ezzine[125]	SVD	Glottal flow parameters	Larynx cancer	ANN	96.9

(*Continued*)

TABLE 1. (*Continued*)

Sl. No.	Author	Database	Feature Extraction Technique	Voice Disorder Type	ML Technique	Best Accuracy (%)
73	Ali et al[53]	King Abdul Aziz University	Voice contour	Cysts, laryngopharyngeal reflux disease, unilateral vocal fold paralysis, polyps, and sulcus vocalis	SVM	100
74	Ali et al[54]	MEEI	MDVP parameters with fractal dimension	Not mentioned	SVM	94.71
75	Benba et al[51]	Local Database	MFCC	Parkinson's disease	SVM	100
76	Benba et al[52]	Local Database	MFCC, perceptual linear prediction, and ReAlitive SpecTrAl PLP	Parkinson's disease and other neurological diseases	Linear SVM	90
77	Chen et al[65]	UCI machine learning repository	Feature reduction using mRMR, Information gain, relief	Parkinson's disease	ELM	95.97
78	Francis et al[126]	MEEI and local dataset	Modified Mellin transform of log spectrum	Polyps, cyst, vocal fold cancer, presbylarynx,, nodules and laryngo pharyngeal reflux disorder	ANN	96.48 95.92
79	Hammami et al[127]	Svd	Pitch and three first formants	Acute laryngitis, adductor spasmodic, vocal fatigue, vocal tremor, vocal fold edema, laryngeal paralysis	SVM	86%
80	Hemmerling et al[71]	SVD	Fundamental frequency, jitter and shimmer coefficients, energy, 0-, 1-, 2-, 3-order moment, kurtosis, power factor, 1-, 2-, and 3-formants amplitude, 1-, 2-, and 3-formants frequency, maximum and minimum value of the signal and 10 MFCC	Hyperfunctional dysphonia, vocal cord paresis and other pathologies mentioned in the database	Random Forest	100
81	Kohler et al[15]	Rio de Janeiro	Glottal flow signal parameters and MFCC	Nodule, paralysis	ANN HMM SVM	96.6 92 97.2
82	Muhammad et al[55]	MEEI SVD	Features extracted from the vocal tract area	Cyst, paralysis, polyp	SVM	99.22 ± 0.01 94.7 ± 0.21

(*Continued*)

TABLE 1. (***Continued***)

Sl. No.	Author	Database	Feature Extraction Technique	Voice Disorder Type	ML Technique	Best Accuracy (%)
83	Shahsavari et al[128]	Parkinson's disease dataset	Practicle swarm optimization based feature selection	Parkinson's disease	ELM	88.72
84	Sharma and Gupta[129]	Local database	ECP (energy between consecutive peaks) and Automatic Speech Recognition (ASR) (average slew rate)	Parkinson's disease	SVM	86
85	Wang et al[130]	People's Liberation Army General Hospital	Multidimensional acoustic features based on *Grade*, Roughness, Breathiness, Asthenia, Strain (GRBAS)	Vocal fold lesions, vocal cord paralysis, arytenoid granuloma, precancerous vocal cord lesions, vocal cord carcinoma and laryngectomized speech	ELM	89.79
86	Ali et al[41]	MEEI	LPC	Variety of voice disorders	GMM	99.94 (sustained vowel) 99.75 ± 0.8 (continuous speech)
87	Al-nasheri et al[8]	SVD MEEI AVPD	MDVP	Cysts, paralysis, polyps	SVM	99.68%, 88.21%, and 72.53% for the three mentioned databases, respectively
88	Al-nasheri et al[56]	MEEI SVD AVPD	Peak and lag values are extracted using correlation functions	Cysts, paralysis, polyps	SVM	99.255, 98.941 , and 95.188 for the three mentioned databases, respectively
89	Amami and Smiti[58]	MEEI	Density-based spatial clustering of applications with noise (DBSCAN) clustering technique to detect noise	Paralysis, keratosis, vocal polyp, adductor	SVM with RBF kernel	98
90	Benba et al[131]	Local database	Features extracted from the time, frequency, and cepstral domains	Parkinson's disease	SVM KNN	88% with KNN
91	Cordeiro et al[57]	MEEI	MFCC, line spectral frequencies	Vocal fold nodule, edema, paralysis	Heirarchical classifier	98.7

(*Continued*)

TABLE 1. (*Continued*)

Sl. No.	Author	Database	Feature Extraction Technique	Voice Disorder Type	ML Technique	Best Accuracy (%)
92	Dahmani and Guerti[132]	SVD	MFCC, jitter, shimmer, and fundamental frequency	Spasmodic dysphonia and paralysis diseases	NBN	90
93	Muhammad et al[133]	MEEI, SVD, and AVPD	Interlaced derivative pattern	Vocal folds cysts, unilateral vocal fold paralysis, and vocal fold polyp	SVM	88.5% for detection and 90.3% for classification
94	Shia and Jayasree[134]	SVD	Discrete wavelet transform	Not mentioned	Feed forward neural network	93.3
95	Teixeira et al[67]	SVD	Jitter, shimmer, and harmonic-to-noise ratio	Not mentioned	ANN	100 for female voices and 90 for male voices
96	Harar et al[135]	SVD	-	Not mentioned	Deep neural networks	71.36
97	Holi[81]	-	MFCC fused with time domain parameters	Neurological voice disorder	Hybrid classifier (GM M and SVM)	94.3
98	Al-nasheri et al[59]	MEEI SVD AVPD	Frequency bands	Vocal fold cysts, unilateral vocal fold paralysis, vocal fold polyps	SVM	99.54 99.53 96.02

Abbreviations: LVQ, learning vector quantization; mRMR, minimum redundancy maximum relevance; RBF, radial basis function; UCI, University of California, Irvine.

SVM

SVM has its roots in statistical learning theory and has shown promising results in many practical applications, which work very well for high-dimensional data.[42] The basic idea here is to construct a hyperplane with a maximum margin that separates the features of different classes by representing the decision boundary is represented as a set of support vectors.

Behroozmand and Almasganj[43] have investigated the role of energy and entropy features, extracted using wavelet packet decomposition for the speech signal with unilateral vocal fold paralysis. The extracted features are optimized using genetic algorithm and classified using SVM with linear kernel. Accuracy of 100% is achieved with entropy features and 93.62% with energy features. Number of active sub-bands obtained for the entropy features is 32, whereas it is 11 for energy features. The highest recognition rate is achieved by using set of 13 active sub-band entropy features. Therefore, the researchers have concluded that the entropy features provided the valuable tool for diagnosis of laryngeal paralysis.

Markaki and Stylianou[44] have applied SVM classifier on set of sustained vowel voice recordings from MEEI database for pathology detection and reported an accuracy of 94.1%. They have briefly analyzed the various signal processing methods for feature extraction from voice and also mentioned the challenges involved in some of the techniques like accurate estimation of fundamental frequency, glottal waveform, etc. They have used a joint acoustic and modulation frequency representation, referred as modulation spectrum, along with dimensionality reduction for detection and classification of voice disorders. These patterns of amplitude modulations are expected to be distorted when voice pathology is present and hence, provide clues for its detection and classification. It is concluded that, even though the results are encouraging, more experiments need to be done on dataset from different hospitals for everyday clinical use of such systems. Saeedi et al[45] proposed a wavelet-based method to discriminate between normal and pathological voices by extracting eight energy features from the wavelet filter banks in which filter banks are constructed using lattice factorization. They have used SVM classifier to classify the voice samples from MEEI database and one private database obtaining 100% classification accuracy for both the databases. Arjmandi and Pooyan[46] have done the comparative analysis of feature extraction techniques like short-time Fourier transform, continuous wavelet transforms, and wavelet packet transform (WPT) with feature reduction techniques, PCA and LDA and SVM classifier. The researchers found that entropy features in the sixth level of WPT decomposition along with LDA and SVM is the most optimum algorithm leading to the recognition rate of 100%. Due to the nonlinearity and randomness behavior of pathological voices, entropy measures lead to a good discrimination of normal and pathological voices as entropy is a statistical measure of randomness. Uloza et al[47] have investigated the effectiveness of different feature sets in the classification of voice recordings using SVM. They have considered pitch and amplitude perturbation measures, frequency features, mel-frequency features, cepstral energy features, MFCCs, autocorrelation features, HNR in spectral domain (HNR spectral), HNR in cepstral domain (HNR cepstral), LP coefficients, and LP cosine transform (LPCT) coefficients. They have classified normal, nodular, and diffuse lesion voices and achieved over 90% accuracy using sequential committee of SVM. The researchers have compared the results of proposed method with three human experts and observed that, using only sustained vowels as information source, the proposed method is significantly better than human experts.

Muhammad and Melhem[48] have done a contribution in their paper, by evaluating MPEG-7 audio low-level features for pathology detection and classification. They have also used an FDR for feature selection. The features are classified using SVM with 99.99% accuracy for pathology detection and up to 100% accuracy for pathology classification among the four diseases, nodules, polyp, keratosis, and adductor together with paralysis and nonparalysis classification. Saidi and Almasganj[49] proposed a method for classification of normal and pathological voices with M-band wavelet as feature extraction technique and linear kernel SVM classifier. They have investigated the optimum four and five band wavelets and achieved 100% accuracy. Orozco- Arroyave et al[50] have considered three types of voice disorders that includes Parkinson's disease (PD), laryngeal pathologies, cleft lip and palate (CLP). Since the characteristics and symptoms of the considered pathology are very different, they have proposed different strategies for modeling these pathologies. The voice signals are analyzed with four different methods including noise content measures, spectral−cepstral modeling, nonlinear features, and the measurements to quantify the stability of the fundamental frequency. The main objective of their work is to advance the interpretation and analysis of different voice pathologies. The classification experiments are performed on six databases using SVM with classification accuracies ranging from 81% to 99%. Based on the experimental results, they have stressed on the importance of understanding the origin and organ/tissues involved in pathology disease for deciding the characterization of voice recordings. For example, the periodicity features could be appropriate for related to the stability of vocal fold vibration and noise content features or spectral and cepstral modeling will be suitable for problems related to hoarseness. It is also mentioned that the proposed methods in the paper have limitations and more experiments can be done using techniques like deep neural networks.

Benba et al[51] have discriminated PD and healthy voices with MFCC features from three types of sustained vowels (/a/, /o/, and /u/). To extract voiceprint, they have compressed the frames of the MFCC by calculating their average value. It is reported that voiceprint is a good parameter for the detection of voice disorder in the context of PD and sustained vowel /u/ contains more discriminative analysis than other types of voice recordings. The proposed method has achieved the 100% accuracy with MLP kernel of SVM. Similarly, in Benba et al,[52] they have discriminated between PD and other neurological disorders by extracting MFCC, perceptual linear prediction, and ReAlitive SpecTrAl PLP from voice samples. They obtained 90% classification accuracy using the first 11 coefficients of the PLP and linear SVM kernels. Ali et al[53] have proposed a novel method based on the voice intensity of a speech signal of continuous speech using SVM classifier. A voice contour is formed by determining the peaks from the speech signals. The area under the voice contour is used to discriminate between normal and disordered subjects. Generally, the area under the voice contour of disordered voices is lower than that of normal voices. The advantage of the proposed feature is that it does not require the estimation of fundamental frequency. They have used the database obtained from the King Abdul Aziz University, which consists of voice recordings of vocal folds cysts, laryngopharyngeal reflux disease, vocal folds polyps, unilateral vocal folds paralysis, and sulcus vocalist patients as well as normal person with a classification accuracy of 100%. Ali et al[54] proposed a multiband approach for the detection of voice disorders suing SVM. This multiband approach is based on a three-level DWT and in each band, fractal dimension (FD) of the power spectrum is estimated. Experiments are done by appending MDVP parameters with FD. They have used MEEI database, which contains 173 pathological and 53 normal voices. Since 5 out of 173 pathological voices did not contain MDVP parameters, they used only 168 pathological voices. The results are expressed in term of sensitivity (SEN), specificity (SPE), accuracy (ACC), and the area under the curve. They observed an improvement of 2.26% in accuracy and 1.45% in area under the curve from their previous work by combining FD of all levels with 22 MDVP parameters.

Muhammad et al[55] developed an automatic voice pathology detection system based on voice production theory, extracting five measurements of the irregularity, namely average, variance, and the ratio of the successive tubes, skewness, and kurtosis from the vocal tract area. The features extracted from the vocal tract area are connected to the glottis. Hence, the features extracted from the pathological voice samples exhibit irregular patterns over frames in case of sustained vowel, which contributes in discrimination of normal and pathological voices. The authors have concluded that supraglottic tract has more contributions than the remaining part of the vocal tract and the variance of vocal tract tubes across the utterance is more important than the mean of those tubes. The extracted features are classified using SVM and evaluated on MEEI and SVD databases. The proposed method has achieved an accuracy of 99.22 ± 0.01 for MEEI database and 94.7 ± 0.21 for SVD. Al-nasheri et al[8] investigated the MDVP parameters in automatic detection of voice pathologies from multiple databases using SVM. They have used three databases such as AVPD, MEEI, and SVD and considered three common voice disorders (cyst, paralysis, and polyp) from these databases. The MDVP parameters extracted using a computerized speech lab program are ranked using FDR technique. A t test is performed to find out any significant differences in means of normal and pathological samples. They have observed a clear difference in the performance of MDVP parameters using three databases. The highly ranked parameters also changed from one database to another and the best accuracies obtained are 99.68%, 88.21%, and 72.53% for the SVD, MEEI, and AVPD, respectively. Al-nasheri et al[56] have performed a voice pathology detection and classification on different frequency regions using correlation functions and SVM classifier. The peak and their corresponding lag values are estimated using correlation functions. These features are experimented on different frequency bands to investigate the contribution of each band on the automatic detection and classification process. It is reported that the frequency bands between 100 and 8000 Hz contributes more in detection and classification of voice disorders. The accuracies of detection and classification varied from one database to another. The best detection rate obtained for MEEI, SVD, and AVPD databases are 99.809%, 90.979%, and 91.168% and that of classification rates are 99.255%, 98.941%, and 95.188%, respectively. Cordeiro et al[57] have designed three-level hierarchical SVM classifier using Gaussian kernel that is trained in a way that first level SVM classify physiological larynx pathologies and healthy voices, second classifier was used to compare neuromuscular larynx pathologies and healthy voices, whereas third model was used to compare physiological larynx pathologies and neuromuscular larynx pathologies. Both sustained vowel and continuous speech database are used as input to extract MFCC and line spectral frequencies features. The hierarchical classifier has achieved an accuracy of 98.7% for voice pathology identification, whereas 81.3% for classification of two classes of unhealthy voices.

Amami and Smiti[58] have proposed incremental Density-based spatial clustering of applications with noise (DBSCAN)-SVM in order to detect noises, analyze, and classify pathological voice from normal voice. The advantages of using DBSCAN algorithm are: it can discover clusters of arbitrary shape, can distinguish noise, uses spatial access methods, and is efficient even for large spatial databases. They used SVM with radial basis function kernel for classification and evaluated the method on MEEI database with an accuracy of 98%. It is concluded that the proposed method has the ability to handle incremental and dynamic voices database, which changes over time. Al-nasheri et al,[59] in their paper, concentrated on developing an accurate and robust feature extraction for detecting and classifying voice pathologies by investigating different frequency bands using autocorrelation and entropy. Maximum peak values and their corresponding lag values have been extracted from each frame of a voiced signal by using autocorrelation, and are used as features. The experiments are carried on MEEI, SVD, and AVPD databases with SVM classifier. They have also performed U tests to investigate the difference between the means of the normal and pathological samples. The variations in the accuracies of both detection and classification based on frequency band and database is reported. It is mentioned that the most contributive bands in both detection and classification were between 1000 and 8000 Hz. The highest accuracies obtained in case of detection were 99.69%, 92.79%, and 99.79% for MEEI, SVD, and AVPD, respectively, and in case of classification were 99.54%, 99.53%, and 96.02% for MEEI, SVD, and AVPD, respectively.

ANN

ANN algorithm works based on the concepts of biological neural networks in human. It consists of set of neurons connected to each other and the output of one neuron can be fed as input to another neuron. The neurons are arranged in layers, ie, input layer, hidden layer, and output layer also referred as multilayer perceptron (MLP) where the number of neurons in each layer and the total numbers of layers depend on the type of applications.

Ritchings et al[60] have proposed a system for assessing the voice quality of patient after laryngeal cancer therapy using MLP. The quality assessment is ranked into seven levels from 0 (least abnormal) to 6 (most abnormal). The database is created with the help of the clinicians by capturing both EGG and acoustic data channels synchronously at 20 kHz for up to 3 seconds while the subject phonated the vowel /i/ as steadily as possible. These recordings are processed to compute long term features like mean of fundamental frequency, the standard deviation of fundamental frequency, the percentage of voiced data and short-term features related to the spectral envelope of the first few glottal harmonics, and the glottal noise. MLP with two layers (20−40 neurons) and seven outputs is designed to train and test the dataset and the highest accuracy reported is 92% with combined long-term and short-term features. Godino-Llorente and Gomez-Vilda[61] have done experiments to detect the laryngeal pathological voices (polyps, nodules, cysts, sulcus, edemas, carcinoma, etc.) on MEEI database using two neural network classification approaches, namely MLP and learning vector quantization with MFCC features. They have reported an outperformance of learning vector quantization method over MLP with 96% classification accuracy. Based on the literature survey and experimental results, they have stressed that MFCC is a good parameterization technique for detecting the voice diseases compared to other feature extraction techniques used in the literature.

An MLP with 64 input nodes and 1 output node is designed and evaluated in Crovato and Schuck[62] for classification of voices into five pathological groups (chronic laryngitis, degenerative, incorrect mobility, organic alterations, and organic growth) and one normal group from the database created at PUCRS Hospital. Wavelet packet coefficients are used for feature computation and the global success rate in the range 87.5% to 100% is reported. Hariharan et al[63] proposed a new feature extraction technique to extract the time domain features. The main objective of developing this feature extraction technique is that it does not require the computation of fundamental frequency, which is very difficult to estimate from the pathological voice samples. These time domain features are used to classify vocal fold paralysis and edema using probabilistic neural network (PNN). The proposed method achieved classification accuracy of 90%. Akbari and Arjmandi[64] have proposed a method using discrete WPT as feature extraction technique, multiclass LDA as dimensionality reduction technique and multilayer neural network as classifier. The proposed system showed average classification accuracy of 96.67% and 97.33% for the structure composed of wavelet packet-based energy and entropy features, respectively.

Chen et al[65] explored the ability of extreme learning machine (ELM) and kernel ELM in early diagnosis of PD from University of California, Irvine ML repository. ELM is a type of ANN, used for single-hidden layer feed-forward neural networks, which was initially proposed by Huang et al.[66] It chooses the hidden nodes randomly and tends to provide good generalization performance at extremely fast learning speed. The dataset used in Chen et al[65] contained some MDVP parameters like two measures of ratio of noise to tonal components in the voice, two nonlinear dynamical complexity measures, signal fractal scaling exponent, and three nonlinear measures of fundamental frequency variation. They have observed an improvement in the performance by reducing the feature space using techniques like minimum redundancy maximum relevance, Information Gain (IG), relief and t test feature selection techniques for ranking the features. The experiments are conducted on various subsets of features. The performance of ELM mainly depends on the hidden neurons and activation function, whereas the performance of KELM depends on the constant parameter C and kernel parameter γ. The authors have experimented these factors to get the optimal results up to 95.97% average accuracy. Teixeira et al[67] evaluated the performance of different features like jitter, shimmer, and HNR in assessment of voice disorders using ANN. Both male and female voice samples are considered from SVD database, an accuracy of 100% for female voices and 90% for male voices is achieved.

K-nearest neighbor classifier

K-nearest neighbor (KNN) is one of the simplest ML algorithms, also called as lazy learner. For a given test data, it uses distance measure to decide the KNNs in the train data. The class labels of the KNNs are analyzed to make the final decision.[68]

KNN algorithm is used in Hadjitodorov and Mitev[69] for pathology detection. Acoustic parameters like shimmer, jitter, several HNRs along with new parameters for estimation of the turbulent noise in voice signals (Turbulent Noise Index) and for the “breathy” voice characterization (Normalized First Harmonic Energy) are used as features. The classification accuracy up to 96.1% is achieved for MEEI database. Similarly, KNN algorithm is used by Shama and Cholayya[70] for detection of laryngeal pathologies like adductor, paralysis, cyst, leukoplakia, vocal fold polyp, polyp degenerative, vocal fold edema, and vocal nodules. Here, they used HNR measure and the critical-band energy spectrum as features. HNRs are estimated at four different frequency bands and used as one set of features. The normalized energies are obtained by filtering the voiced speech signals using 21 critical bandpass filters, which will mimic the human auditory neurons and used as another set of features. The set HNR features achieved 94.28% accuracy, whereas critical energy spectrum features achieved 92.38% accuracy with KNN classifier. The results obtained have shown that these features can be used as a tool to supplement the perceptual evaluation of speech for the detection of suspected laryngeal pathologies. This method requires only shorter length of speech data for the analysis and computationally less expensive as compared to the extraction of fundamental frequency and noise measures.

Decision trees

A decision tree is a tree whose internal nodes are tests (on input patterns) and whose leaf nodes are categories (of patterns).[68] It is constructed based on the information given in the training set data. Hemmerling et al[71] have evaluated different methods of speech signal analysis for the detection of voice pathologies using the Random Forest (RF) classification algorithm. The idea behind RF is to combine many binary decision trees, which are built using different bootstrap samples of the original data and random subsets to obtain an accurate predictor. The advantage of the RF method is that it is robust against overfitting as more trees are added, and, it produces a limiting value of the generalization error. In this work, they have extracted the 28 acoustic features such as fundamental frequency; jitter and shimmer coefficients; energy; 0-, 1-, 2-, and 3-order moment; kurtosis; power factor; 1-, 2-, and 3-formants amplitude; 1-, 2-, and 3-formants frequency; maximum and minimum value of the signal; and 10 MFCC from the SVD database voice samples. The proposed method has achieved 100% classification accuracy.

Linear classifier

The main contribution of Kaleem et al[72] is presenting feature extraction technique based on empirical mode decomposition using which one can extract meaningful information from both time and frequency domains from the continuous speech. The advantage of using continuous speech is that it does not require the separation of sustained vowel from the speech, which is very erroneous task in case of pathological voices. The extracted features are classified using linear classifier and achieved 95.7% classification accuracy.

K-means clustering

K-means clustering is an unsupervised ML technique[68] in which the objects are grouped together based on the Euclidean distance measure. Each cluster is represented by a cluster mean and new object is assigned to a particular cluster based on its distance from cluster mean. Ghoraani and Krishnan[73] have used K-means clustering and proposed a novel methodology for automatic classification of pathological voices by computing features using adaptive time−frequency distribution and non-negative matrix factorization applied on MEEI database. In this paper, they have mentioned the shortcomings of the feature extraction approaches, which require the signal to be segmented into short frames as the pathological voices are nonstationary and segmenting it at nonstationary part may lead to loss of important information. To overcome those limitations, they proposed a new approach to extract the TF features from the speech so that it can capture the dynamic changes in the pathological speech. The extracted features are classified using K-mean clustering technique with 98.6% classification accuracy.

Combined classifiers

To achieve improvement in classification accuracy, different classifier can be combined together referred as multiple classifier system.[74,68] It can be designed by combining the techniques involved in different classifiers into a system called *ensemble design* or by developing a module called fuser, which combines the output labels of different classifiers. In this section, along with ensemble classifier, we also discuss the research works that has used more than one classifier for comparison.

Gelzinis et al[75] have investigated the performance of 11 features namely pitch and amplitude perturbation measures, frequency features, mel-frequency features, cepstral energy features, MFCCs, autocorrelation features, HNR spectral, HNR cepstral, linear prediction coefficients, LPCT coefficients, and feature set used in the commercial "Dr. Speech" software. They have studied two types of classification tasks like two-class classification (healthy and pathological) and three-class classification (one healthy and two pathological —diffuse and nodular). The studies were made using the mixed gender database containing voice recordings of the sustained phonation of the vowel sound /a/. A KNN classifier, SVM, and a committee of classifiers were used to perform the classification tasks. It is concluded that the pitch and amplitude perturbation measures provided the best overall performance in both classification tasks. They also observed that the cepstral energy, autocorrelation features, and the LPCT coefficients were less effective in two-class classification task, whereas mel-frequency features were less effective in the three-class classification. On an average, single KNN classifier outperformed the SVM classifier. However, the highest classification accuracy of 95.5% is achieved by using the SVM committee.

Das[76] compared the performances of different classification systems like neural networks, DMneural, regression, and decision tree in detecting PD. They have used the PD dataset that comprises biomedical voice measures which will characterize healthy persons and patients. The authors have used variable selection component to select the best features in order to reduce the dimensionality of the feature vector. The highest classification rate of

92.9% is achieved using neural network. Arias-Londono et al[77] have highlighted that even though MFCC is a popular feature, it is not able to characterize the nonlinear process involved in voice production. Hence, they proposed a new approach for speech feature extraction using nonlinear analysis of time series for automatic classification of normal and pathological voices. The set of 11 complexity measures are combined with MFCC (combined with noise parameters) features for better representation of voice data. A fusion of generative approach based on GMM, and discriminative approach based on SVM is used for classification. The features are fed into two different GMM for training. The testing is done with GMM at one level and later fed into SVM for second level of classification. This approach is applied on MEEI database with a classification accuracy of 98.23%. Arjmandi et al[78] performed classification experiment on randomly selected 50 normal and 50 pathological voice samples from MEEI database. They extracted 33 acoustic parameters and used only 22 parameters as some of them did not reflect voice quality. They investigated two feature reduction techniques like PCA and LDA and feature selection techniques like individual, forward, backward, and branch-and-bound. To evaluate the feature vectors, authors have used six classifiers such as quadratic discriminant classifier, nearest mean classifier, Parzen classifier, KNN, support vector classifier, and neural network. After evaluating the different combinations of feature reduction/feature selection methods and classifiers, they have proposed a combined scheme of LDA and SVM with recognition rate of 94.26%. Among feature selection methods, the individual technique has achieved best accuracy of 91.55% with SVM.

Hariharan et al[79] proposed a new hybrid intelligent system, which includes feature preprocessing techniques, feature reduction techniques, and classifiers to classify PD patients form healthy people. They have used PD dataset from University of California, Irvine ML database. Model-based clustering (GMM) is used as feature preprocessing technique, whereas PCA, LDA, sequential forward selection, and sequential backward selection are used as feature reduction/selection techniques. Three classifiers, namely, least-square SVM, PNN, and general regression neural network are used to evaluate the performance of features. This hybrid system achieved 100% accuracy with combination of GMM based feature weighting, LDA/sequential backward selection, and least-square SVM/PNN/general regression neural network. Jothilakshmi[80] has used database created with the help of Rajah Muthiah Medical College Hospital, Annamalai University, which consists of voice samples of patients with cerebral palsy, dysarthria, hearing impairments, laryngectom, mental retardation, left-side paralysis, quadriparesis, stammering, stroke, and tumour in vocal tract. In this work, the author has developed a two-level model using HMM and GMM to detect the type of voice pathology. Initially, the voice sample will be classified as either normal or pathological and in second level, it will be classified into one of the voice disorder listed above if the voice is pathological considered in the database. The 12 LPC and 13 MFCC along with its first and second derivatives are used as features. The obtained results have shown that GMM is well suited for voice pathology classification (95.74%) and HMM for pathological voice detection (94.44%).

Kohler et al[15] have used glottal flow parameters of the vocal folds to classify vocal fold nodules and vocal fold paralysis. The parameters of the glottal signals can be obtained through inverse filtering method. Time domain and frequency domain parameters are extracted from the glottal signals and used in various experiments. They have used a database composed of voice recordings of patients with nodules, and paralysis as well as normal voices of both men and women. This database was obtained from a speech therapist in Rio de Janeiro. The authors have used ANN, HMM, and SVM for classification. From the obtained results, it is concluded that glottal signal parameters are more relevant to discriminate pathologies of the vocal folds than MFCC. The highest accuracy achieved is 97.2% using SVM. Holi[81] have developed a hybrid classifier for detection of neurological disorders. The acoustic characteristics of a voice signals with neurological disorders may change due to incomplete closure of glottis. These changes may help in detecting the certain neurological diseases. The wavelet transform and MFCCs are fused with the time-domain features and fed into a hybrid model designed using GMM and SVM. LP-coded MFCCs computed for selected six-level DWT are given to GMM. The output of which is combined with SVM scores obtained with time-domain features as input and given to another SVM, which makes the decision to classify the data as normal or neurological-disordered voice. The proposed system achieved classification accuracy of 94.3%.

DISCUSSIONS

Various voice disorders, databases, feature extraction, and ML techniques used in AVDD research so far have been discussed in this paper. It is observed that SVM classifier is used most widely compared to any other ML techniques. It may be due to the fact that SVM classifier performs better for high-dimensional data as well as small sized dataset and it is quite difficult to get large database of pathological voices for the research. In contrast, ANN requires a large dataset for training the model to give better classification accuracy.

Most of the innovations or novel approaches proposed in the literature are mainly focusing on the voice characterization methods, which is referred to it as feature extraction technique. Many authors have claimed to propose new method for feature extraction and have reported an improvement in performance. The feature extraction step can be skipped by using deep learning techniques, which extracts features inherently but needs large training examples. Even though Deep Neural Network (DNNs) have emerged as a powerful ML technique, there are very few works done using deep learning, which may be due to the nonavailability of large databases.

It is found that most of the papers have reported high accuracy in pathology detection, which has gone up to 100% in few cases. However, the results are specific to the database used and it cannot be guaranteed that the system developed for one database can be applied to other databases with same degree of performance.

To carry out research work in this area and to develop real-time clinical system, the availability of large standard databases is essential. To our knowledge, there are limited numbers of standard benchmark databases, which can be used for such research since the creation of a database requires the collaboration of clinicians who are experts in the field of vocal disorder diagnosis. Using the local database, an automated system can be developed for a particular hospital. Generally, clinicians are using audio files for diagnosis, but in long term they may not keep track or maintain the audio files based on the disease rectified.

CONCLUSIONS

In this paper, we have analyzed cardinal research work related to AVDD system. Organization of the contents in this paper is based on the type of voice disorders, feature extraction, and ML techniques that have been used. We have discussed the theoretical aspects of voice disorders, feature extraction techniques, ML techniques, and reviewed the performance of some of the significant research works done in this area. The research work experiments are done by selecting the subjects belonging to a few of the diseases randomly chosen from the standard database. The voices are generally categorized as normal and pathological in most of the papers. However, Alzheimer's disease and PD have been specifically targeted in some of the studies.

Research work in this area has a wide scope due to its applicability for society. With this concern there is a lot of scope in developing more standard databases, inventing new features, applying DNN, and working with a specific voice disorder. We believe that this paper not only gives an insight into the type of research that has been done over the years, but also helps highlight the areas that need further experimentation and analysis. We wish to further our study in finding the ideal AVDD system, which is accurate, efficient, and adequate, to make early detection and diagnosis of the various voice disorders as quick and effortless as possible.

Acknowledgments

We thank Nitte University, Mangalore, for providing the support in terms of resources for the research work described in this paper.

SUPPLEMENTARY DATA

Supplementary data related to this article can be found, online at doi:10.1016/j.jvoice.2018.07.014.

REFERENCES

1. Titze IR, Verdolini K. *Vocology: The Science and Practice of Voice Habilitation*. Salt Lake City, UT: National Center for Voice and Speech; 2012.
2. Dejonckere P, Bradeley P, Clemente P, et al. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *Eur Arch Otorhinolaryngol*. 2001;258:77–82.
3. NIDCD. *2012-2016 Strategic Plan,*Vol. 2.Bethesda, MD: Bethesda, MD: National Institute on Deafness and Other Communication Disorders (NIDCD), U.S. Department of Health and Human Services; 2012.
4. American Speech-Language-Hearing Association and Others. "Council for clinical certification in audiology and speech-language pathology," Retrieved September, vol. 15, 2015.
5. Verdolini K, Rosen CA, Branski RC, et al. *Classification Manual for Voice Disorders—I. Special Interest Division 3: Voice and Voice Disorders*. Lawrence Erlbaum Associates, Inc; 2006.
6. Titze IR, Švec JG, Popolo PS. Vocal dose measures: quantifying accumulated vibration exposure in vocal fold tissues. *J Speech Lang Hear Res*. 2003;46:919–932. http://dx.doi.org/10.1044/1092-4388(2003/072).
7. Boominathan P, Samuel J, Arunachalam R, et al. Multiparametric voice assessment: Sri Ramachandra University protocol. *Indian J Otolaryngol Head Neck Surg*. 2014;66:246–251.
8. Al-nasheri A, Muhammad G, Alsulaiman M, et al. An investigation of multidimensional voice program parameters in three different databases for voice pathology detection and classification. *J Voice*. 2017;31. 113.e9−113.e18.
9. Kasuya H, Ogawa S, Kikuchi Y, et al. An acoustic analysis of pathological voice and its application to the evaluation of laryngeal pathology. *Speech Commun*. 1986;5:171–181.
10. Sonu, Sharma RK. Disease detection using analysis of voice parameters. *Int J Comput Sci Commun Technol*. 2012;4:6–10.
11. Al-nasheri A, Ali Z, Muhammad G, et al. Voice pathology detection with MDVP parameters using Arabic voice pathology database. *IEEE 5th National Symposium on Information Technology: Towards New Smart World (NSITNSW)*. 1–5.
12. Rabiner LR, Juang B-H. *Fundamentals of Speech Recognition*. Prentice Hall; 1993.
13. Slaney M. *Toolbox: a Matlab toolbox for auditory modeling. Work Technical Report,*. Interval Research Corporation; 1998:29–32.
14. Atal BS, Hanauer SL. Speech analysis and synthesis by linear prediction of the speech wave. *J Acoust Soc Am*. 1971;50:637–655.
15. Kohler M, Vellasco MM, Cataldo E. Analysis and classification of voice pathologies using glottal signal parameters. *J Voice*. 2016;30:549–556.
16. Kantardzic M. *Data Reduction*. John Wiley & Sons, Inc; 2003:53–86.
17. Deller JR, Anderson DJ. Automatic classification of laryngeal dysfunction using the roots of the digital inverse filter. *IEEE Trans Biomed Eng*. 1980;12:714–721.
18. Childers DG, Bae KS. Detection of laryngeal function using speech and electroglottographic data. *IEEE Trans Biomed Eng*. 1992;39:19–25.
19. Cairns DA, Hansen JH, Riski JE. A noninvasive technique for detecting hypernasal speech using a nonlinear operator. *IEEE Trans Biomed Eng*. 1996;43:35.
20. Accardo AP, Mumolo E. An algorithm for the automatic differentiation between the speech of normals and patients with Friedreich's ataxia based on the short-time fractal dimension. *Comput Biol Med*. 1998;28:75–89.
21. Parsa V, Jamieson DG. Identification of pathological voices using glottal noise measures. *J Speech Lang Hear Res*. 2000;43:469–485.
22. Hadjitodorov S, Boyanov B, Teston B. Laryngeal pathology detection by means of class- specific neural maps. *IEEE Trans Inf Technol Biomed*. 2000;4:68–73.
23. de Oliveira Rosa M, Pereira JC, Grellet M. Adaptive estimation of residue signal for voice pathology diagnosis. *IEEE Trans Biomed Eng*. 2000;47:96–104.

24. Watts CR, Clark R, Early S. Acoustic measures of phonatory improvement secondary to treatment by oral corticosteroids in a professional singer: a case report. *J Voice*. 2001;15:115–121.
25. Guido RC, Pereira JC, Fonseca E, et al. Trying different wavelets on the search for voice disorders sorting. In: *Proceedings of the Thirty-Seventh Southeastern Symposium on System Theory, 2005 (SSST'05)*. IEEE; 2005:495–499.
26. Umapathy K, Krishnan S, Parsa V, et al. Discrimination of pathological voices using a time-frequency approach. *IEEE Trans Biomed Eng*. 2005;52:421–430.
27. Zhang Y, Jiang JJ, Biazzo L, et al. Perturbation and nonlinear dynamic analyses of voices from patients with unilateral laryngeal paralysis. *J Voice*. 2005;19:519–528.
28. Neto BGA, Fechine JM, Costa SC, et al. Feature estimation for vocal fold edema detection using short-term cepstral analysis. In: *Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering, 2007 (BIBE 2007)*. IEEE; 2007:1158–1162.
29. Gómez-Vilda P, Fernández-Baillo R, Nieto A, et al. Evaluation of voice pathology based on the estimation of vocal fold biomechanical parameters. *J Voice*. 2007;21:450–476.
30. Gómez-Vilda P, Fernández-Baillo R, Rodellar-Biarge V, et al. Glottal source biometrical signature for voice pathology detection. *Speech Commun*. 2009;51:759–781.
31. Zhang Y, Jiang JJ. Acoustic analyses of sustained and running voices from patients with laryngeal pathologies. *J Voice*. 2008;22:1–9.
32. Fontes AI, Souza PT, Neto AD, et al. Classification system of pathological voices using correntropy. *Math Probl Eng*. 2014;2014:1–7.
33. Levinson SE, Rabiner LR, Sondhi MM. An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *Bell Syst Tech J*. 1983;62:1035–1074.
34. Poritz AB. Hidden Markov models: a guided tour. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 7–13.
35. Rabiner LR, Juang BH. Speech recognition: statistical methods. *Encyclopedia of Language & Linguistics*2nd ed. 1–18.
36. Rao PVS. VOICE: an integrated speech recognition synthesis system for the Hindi language. *Speech Commun*. 1993;13:197–205.
37. Gavidia-Ceballos L, Hansen JH. Direct speech feature estimation using an iterative EM algorithm for vocal fold pathology detection. *IEEE Trans Biomed Eng*. 1996;43:373–383.
38. Arias-Londoño JD, Godino-Llorente JI, Sáenz-Lechón N, et al. An improved method for voice pathology detection by means of a HMM-based feature space transformation. *Pattern Recognit*. 2010;43:3100–3112.
39. Muhammad G, Mesallam TA, Malki KH, et al. Multidirectional regression (MDR)-based features for automatic voice disorder detection. *J Voice*. 2012;26. 817.e19−817.e27.
40. Ali Z, Elamvazuthi I, Alsulaiman M, et al. Automatic voice pathology detection with running speech by using estimation of auditory spectrum and cepstral coefficients based on the all-pole model. *J Voice*. 2015;30(6), 757−e7.
41. Ali Z, Muhammad G, Alhamid MF. An automatic health monitoring system for patients suffering from voice complications in smart cities. *IEEE Access*. 2017;5:3900–3908.
42. Vapnik V. *The Nature of Statistical Learning Theory*. Springer Science & Business Media; 2013.
43. Behroozmand R, Almasganj F. Optimal selection of wavelet-packet-based features using genetic algorithm in pathological assessment of patients' speech signal with unilateral vocal fold paralysis. *Comput Biol Med*. 2007;37:474–485.
44. Markaki M, Stylianou Y. Voice pathology detection and discrimination based on modulation spectral features. *IEEE Trans Audio Speech Lang Process*. 2011;19:1938–1948.
45. Saeedi NE, Almasganj F, Torabinejad F. Support vector wavelet adaptation for pathological voice assessment. *Comput Biol Med*. 2011;41:822–828.
46. Arjmandi MK, Pooyan M. An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine. *Biomed Signal Process Control*. 2012;7:3–19.
47. Uloza V, Verikas A, Bacauskiene M, et al. Categorizing normal and pathological voices: automated and perceptual categorization. *J Voice*. 2011;25:700–708.
48. Muhammad G, Melhem M. Pathological voice detection and binary classification using MPEG-7 audio features. *Biomed Signal Process Control*. 2014;11:1–9.
49. Saidi P, Almasganj F. Voice disorder signal classification using m-band wavelets and support vector machine. *Circuits Syst Signal Process*. 2015;34:2727–2738.
50. Orozco-Arroyave JR, Belalcazar-Bolanos EA, Arias-Londoño JD, et al. Characterization methods for the detection of multiple voice disorders: neurological, functional, and laryngeal diseases. *IEEE J Biomed Health Inf*. 2015;19:1820–1828.
51. Benba A, Jilbab A, Hammouch A. Analysis of multiple types of voice recordings in cepstral domain using MFCC for discriminating between patients with Parkinson's disease and healthy people. *Int J Speech Technol*. 2016;19:449–456.
52. Benba A, Jilbab A, Hammouch A. Discriminating between patients with Parkinson's and neurological diseases using cepstral analysis. *IEEE Trans Neural Syst Rehabil Eng*. 2016;24:1100–1108.
53. Ali Z, Alsulaiman M, Elamvazuthi I, et al. Voice pathology detection based on the modified voice contour and SVM. *Biol Inspired Cognit Archit*. 2016;15:10–18.
54. Ali Z, Elamvazuthi I, Alsulaiman M, et al. Detection of voice pathology using fractal dimension in a multiresolution analysis of normal and disordered speech signals. *J Med Syst*. 2016;40:20.
55. Muhammad G, Altuwaijri G, Alsulaiman M, et al. Automatic voice pathology detection and classification using vocal tract area irregularity. *Biocybern Biomed Eng*. 2016;36:309–317.
56. Al-nasheri A, Muhammad G, Alsulaiman M, et al. Investigation of voice pathology detection and classification on different frequency regions using correlation functions. *J Voice*. 2017;31:3–15.
57. Cordeiro H, Fonseca J, Guimarães I, et al. Hierarchical classification and system combination for automatically identifying physiological and neuromuscular laryngeal pathologies. *J Voice*. 2017;31. 384.e9−384.e14.
58. Amami R, Smiti A. An incremental method combining density clustering and support vector machines for voice pathology detection. *Comput Electr Eng*. 2017;57:257–265.
59. Al-Nasheri A, Muhammad G, Alsulaiman M, et al. Voice pathology detection and classification using auto- correlation and entropy features in different frequency regions. *IEEE Access*. 2018;6:6961–6974.
60. Ritchings RT, McGillion M, Moore CJ. Pathological voice quality assessment using artificial neural networks. *Med Eng Phys*. 2002;24:561–564.
61. Godino-Llorente JI, Gomez-Vilda P. Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors. *IEEE Trans Biomed Eng*. 2004;51:380–384.
62. Crovato CDP, Schuck A. The use of wavelet packet transform and artificial neural networks in analysis and classification of dysphonic voices. *IEEE Trans Biomed Eng*. 2007;54:1898–1900.
63. Hariharan M, Paulraj MP, Yaacob S. Detection of vocal fold paralysis and edema using time-domain features and probabilistic neural network. *Int J Biomed Eng Technol*. 2011;6:46–57.
64. Akbari A, Arjmandi MK. An efficient voice pathology classification scheme based on applying multi-layer linear discriminant analysis to wavelet packet-based features. *Biomed Signal Process Control*. 2014;10:209–223.
65. Chen HL, Wang G, Ma C, et al. An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson′ s disease. *Neurocomputing*. 2016;184:131–144.
66. Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. *Neurocomputing*. 2006;70:489–501.
67. Teixeira JP, Fernandes PO, Alves N. Vocal acoustic analysis—classification of dysphonic voices with artificial neural networks. *Proc Comput Sci*. 2017;121:19–26.

68. Alpaydin E. *Introduction to Machine Learning*. MIT Press; 2014.
69. Hadjitodorov S, Mitev P. A computer system for acoustic analysis of pathological voices and laryngeal diseases screening. *Med Eng Phys*. 2002;24:419–429.
70. Shama K, Cholayya NU. Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology. *EURASIP J Appl Signal Process*. 2007;2007:50.
71. Hemmerling D, Skalski A, Gajda J. Voice data mining for laryngeal pathology assessment. *Comput Biol Med*. 2016;69:270–276.
72. Kaleem M, Ghoraani B, Guergachi A, Krishnan S. Pathological speech signal analysis and classification using empirical mode decomposition. *Med Biol Eng Comput*. 2013;51:811–821.
73. Ghoraani B, Krishnan S. A joint time-frequency and matrix decomposition feature extraction methodology for pathological voice classification. *EURASIP J Adv Signal Process*. 2009;2009 928974.
74. Woźniak M, Graña M, Corchado E. A survey of multiple classifier systems as hybrid systems. *Inf Fusion*. 2014;16:3–17.
75. Gelzinis A, Verikas A, Bacauskiene M. Automated speech analysis applied to laryngeal disease categorization. *Comput Methods Prog Biomed*. 2008;91:36–47.
76. Das R. A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Syst Appl*. 2010;37:1568–1572.
77. Arias-Londono JD, Godino-Llorente JI, Sáenz-Lechón N, et al. Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients. *IEEE Trans Biomed Eng*. 2011;58:370–379.
78. Arjmandi MK, Pooyan M, Mikaili M, et al. Identification of voice disorders using long-time features and support vector machine with different feature reduction methods. *J Voice*. 2011;25:e275–e289.
79. Hariharan M, Polat K, Sindhu R. A new hybrid intelligent system for accurate detection of Parkinson's disease. *Comput Methods Prog Biomed*. 2014;113:904–913.
80. Jothilakshmi S. Automatic system to detect the type of voice pathology. *Appl Soft Comput*. 2014;21:244–249.
81. Holi MS. Wavelet transform features to hybrid classifier for detection of neurological-disordered voices. *J Clin Eng*. 2017;42:89–98.
82. Weeks M. *Digital Signal Processing Using Matlab and Wavelets*. Infinity Science Press LLC; 2006.
83. Martinez CE, Rufiner HL. Acoustic analysis of speech for detection of laryngeal pathologies. In: *Proceedings of 22nd Annual IEEE International Conference on Engineering in Medicine and Biology Society*. 3, IEEE; 2000:2369–2372.
84. Dibazar AA, Narayanan S, Berger TW. Feature analysis for automatic detection of pathological speech. In: *Proceedings of the Second Joint Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference*. 1, IEEE; 2002:182–183.
85. Nayak J, Bhat PS. Identification of voice disorders using speech samples. *TENCON 2003. Conference on Convergent Technologies for the Asia-Pacific Region*. 3, IEEE; 2003:951–953.
86. Ananthakrishna T, Shama K, Niranjan UC. k-means nearest neighbor classifier for voice pathology. In: *Proceedings of the IEEE INDICON 2004 First India Annual Conference*. IEEE; 2004:352–354.
87. Behroozmand R, Almasganj F. Comparison of neural networks and support vector machines applied to optimized features extracted from patients' speech signal for classification of vocal fold inflammation. In: *Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology*. IEEE; 2005:844–849.
88. Fonseca ES, Guido RC, Silvestre AC, et al. Discrete wavelet transform and support vector machine applied to pathological voice signals identification. *Seventh IEEE International Symposium on Multimedia*. IEEE; 2005:5.
89. Godino-Llorente JI, Gómez-Vilda P, Sáenz-Lechón N, et al. Discriminative methods for the detection of voice disorders. *ISCA Tutorial and Research Workshop (ITRW) on Non-Linear Speech Processing*. .
90. Nayak J, Bhat PS, Acharya R, et al. Classification and analysis of speech abnormalities. *ITBM-RBM*. 2005;26:319–327.
91. Moran RJ, Reilly RB, de Chazal P, et al. Telephony-based voice pathology assessment using automated speech analysis. *IEEE Trans Biomed Eng*. 2006;53:468–477.
92. Schlotthauer G, Torres ME, Jackson-Menaldi MC. Automatic diagnosis of pathological voices. *WSEAS Trans Signal Process*. 2006;2:1260–1267.
93. Fonseca ES, Guido RC, Scalassara PR, et al. Wavelet time- frequency analysis and least squares support vector machines for the identification of voice disorders. *Comput Biol Med*. 2007;37:571–578.
94. Kukharchik P, Martynov D, Kheidorov I, et al. Vocal fold pathology detection using modified wavelet-like features and support vector machines. *15th European on Signal Processing Conference*. IEEE; 2007:2214–2218.
95. Aguiar Neto BG, Costa SC, Fechine JM. LPC modelling and cepstral analysis applied to vocal fold pathology detection. *Int J Funct Inf Personal Med*. 2008;1:156–170.
96. Linder R, Albers AE, Hess M, et al. Artificial neural network- based classification to screen for dysphonia using psychoacoustic scaling of acoustic voice features. *J voice*. 2008;22:155–163.
97. Murugesapandian P, Yaacob S, Hariharan M. Feature extraction based on mel-scaled wavelet packet transform for the diagnosis of voice disorders. *4th Kuala Lumpur International Conference on Biomedical Engineering*. Berlin, Heidelberg: Springer; 2008:790–793.
98. Salhi L, Talbi M, Cherif A. Voice disorders identification using hybrid approach: wavelet analysis and multilayer neural networks. *World Acad Sci Eng Technol*. 2008;45:330–339.
99. Hariharan M, Paulraj MP, Yaacob S. Identification of vocal fold pathology basedonmelfrequencyband energy coefficientsand singular valuedcomposition. *2009 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. IEEE; 2009:514–517.
100. Kotropoulos C, Arce GR. Linear classifier with reject option for the detection of vocal fold paralysis and vocal fold edema. *EURASIP J Adv Signal Process*. 2009;2009:11.
101. Markaki M, Stylianou Y. Normalized modulation spectral features for cross-database voice pathology detection. *Tenth Annual Conference of the International Speech Communication Association*. .
102. Markaki M, Stylianou Y. Using modulation spectra for voice pathology detection and classification. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009 (EMBC 2009)*. IEEE; 2009:2514–2517.
103. Markaki M, Stylianou Y, Arias-Londoño JD, et al. Dysphonia detection based on modulation spectral features and cepstral coefficients. *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. IEEE; 2010:5162–5165.
104. Carvalho RTS, Cavalcante CC, Cortez PC. Wavelet transform and artificial neural networks applied to voice disorders identification. *2011 Third World Congress on Nature and Biologically Inspired Computing (NaBIC)*. IEEE; 2011:371–376.
105. de Bruijn M, ten Bosch L, Kuik DJ, et al. Artificial neural network analysis to assess hypernasality in patients treated for oral or oropharyngeal cancer. *Logoped Phoniatr Vocol*. 2011;36:168–174.
106. Lee J, Jeong S, Hahn M, et al. An efficient approach using HOS-based parameters in the LPC residual domain to classify breathy and rough voices. *Biomed Signal Process Control*. 2011;6:186–196.
107. Tsanas A, Little MA, McSharry PE, et al. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Trans Biomed Eng*. 2012;59:1264–1271.
108. Ali Z, Alsulaiman M, Muhammad G, et al. Vocal fold disorder detection based on continuous speech by using MFCC and GMM. *2013 7th IEEE GCC Conference and Exhibition (GCC)*. IEEE; 2013:292–297.
109. Alsulaiman M, Muhammad G, Ali Z. Classification of vocal fold diseases using RASTA-PLP. In: *Proceedings of the International Conference on Bioinformatics & Computational Biology (BIOCOMP)*. The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp); 2013:1.
110. Belalcazar-Bolanos EA, Orozco-Arroyave JR, Arias-Londono JD, et al. Automatic detection of Parkinson's disease using noise measures

of speech. *IEEE* XVIII Symposium of Image, Signal Processing, and Artificial Vision (STSIVA). 1–5.
111. Holi MS. Automatic detection of neurological disordered voices using mel cepstral coefficients and neural networks. *Point-of-Care Healthcare Technologies (PHT), 2013 IEEE*. IEEE; 2013:76–79.
112. Majidnezhad, V., & Kheidorov, I. An ANN-based method for detecting vocal fold pathology. *arXiv preprint arXiv:1302.1772*;2013.
113. Saldanha JC, Ananthakrishna T, Pinto R. Vocal fold pathology assessment using PCA and LDA. In: *Proceedings of 2013 International Conference on Intelligent Systems and Signal Processing (ISSP)*. IEEE; 2013:140–144.
114. Vikram CM, Umarani K. Phoneme independent pathological voice detection using wavelet based MFCCs, GMM-SVM hybrid classifier. In: *Proceedings of* 2013 International IEEE Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE; 2013:929–934.
115. Cordeiro H, Fonseca J, Meneses C. Spectral envelope and periodic component in classification trees for pathological voice diagnostic. *IEEE 36th Annual International Conference on Engineering in Medicine and Biology Society (EMBC)*. 4607–4610.
116. Fezari M, Amara F, El-Emary IM. Acoustic analysis for detection of voice disorders using adaptive features and classifiers. In: *Proceeeings of 2014th International Conference on Circuits, Systems and Control, Switzerland, February 22−24, 2014*. .
117. El Emary IMM, Fezari M, Amara F. Towards developing a voice pathologies detection system. *J Commun Technol Electron*. 2014;59:1280–1288.
118. Novotný M, Rusz J, Čmejla R, et al. Automatic evaluation of articulatory disorders in Parkinson's disease. *IEEE/ACM Trans Audio Speech Lang Process*. 2014;22:1366–1378.
119. Cordeiro H, Fonseca J, Guimarães I, et al. Voice pathologies identification speech signals, features and classifiers evaluation. *Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*. IEEE; 2015:81–86.
120. López-de-Ipina K, Solé-Casals J, Eguiraun H, et al. Feature selection for spontaneous speech analysis to aid in Alzheimer's disease diagnosis: a fractal dimension approach. *Comput Speech Lang*. 2015;30:43–60.
121. Rani KU, Holi MS. GMM classifier for identification of neurological disordered voices using MFCC features. *IOSR J VLSI Signal Process*. 2015;4:44–51.
122. Salehi P. Using patient's speech signal for vocal ford disorders detection based on lifting scheme. *2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI)*. IEEE; 2015:561–568.
123. Vásquez-Correa JC, Arias-Vergara T, Orozco-Arroyave JR, et al. Automatic detection of Parkinson's disease from continuous speech recorded in non-controlled noise conditions. In: *Proceedings of* Sixteenth Annual Conference of the International Speech Communication Association. .
124. Agarwal A, Chandrayan S, Sahu SS. Prediction of Parkinson's disease using speech signal with extreme learning machine. In: *Proceedings of* IEEE International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT). 3776–3779.
125. Aicha AB, Ezzine K. Cancer larynx detection using glottal flow parameters and statistical tools. *International Symposium on Signal, Image, Video and Communications (ISIVC)*. IEEE; 2016:65–70.
126. Francis CR, Nair VV, Radhika S. A scale invariant technique for detection of voice disorders using modified Mellin transform. *IEEE* International Conference on Emerging Technological Trends (ICETT). 1–6.
127. Hammami I, Salhi L, Labidi S. Pathological voices detection using support vector machine. *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*. IEEE; 2016:662–666.
128. Shahsavari MK, Rashidi H, Bakhsh HR. Efficient classification of Parkinson's disease using extreme learning machine and hybrid particle swarm optimization. *IEEE 4th International Conference on Control, Instrumentation, and Automation (ICCIA)*. 148–154.
129. Sharma RK, Gupta AK. Processing and analysis of human voice for assessment of Parkinson disease. *J Med Imaging Health Inf*. 2016;6:63–70.
130. Wang Z, Yu P, Yan N, et al. Automatic assessment of pathological voice quality using multidimensional acoustic analysis based on the GRBAS scale. *J Signal Process Syst*. 2016;82:241–251.
131. Benba A, Jilbab A, Hammouch A. Voice assessments for detecting patients with neurological diseases using PCA and NPCA. *Int J Speech Technol*. 2017;20:673–683.
132. Dahmani M, Guerti M. Vocal folds pathologies classification using Naïve Bayes Networks. *2017 6th International Conference on Systems and Control (ICSC)*. IEEE; 2017:426–432.
133. Muhammad G, Alsulaiman M, Ali Z, et al. Voice pathology detection using interlaced derivative pattern on glottal source excitation. *Biomed Signal Process Control*. 2017;31:156–164.
134. Shia SE, Jayasree T. Detection of pathological voices using discrete wavelet transform and artificial neural networks. *2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*. IEEE; 2017, March:1–6.
135. Harar P, Alonso-Hernandezy JB, Mekyska J, et al. Voice pathology detection using deep learning: a preliminary study. *IEEE* International Conference and Workshop on Bioinspired Intelligence (IWOBI). 1–4.