Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

Research article

Meta-analysis of voice disorders databases and applied machine learning techniques

Sidra Abid Syed [1],*, Munaf Rashid [2] and Samreen Hussain [3]

[1] Biomedical Engineering Department, Ziauddin University Faculty of Engineering Science Technology and Management, Karachi, Pakistan
[2] Software Engineering Department, Ziauddin University Faculty of Engineering Science Technology and Management, Karachi, Pakistan
[3] Vice Chancellor, Begum Nusrat Bhutto Women University, Sukkur, Pakistan

* **Correspondence:** Email: sidra.agha@zu.edu.pk.

Abstract: *Background and Objective:* Voice disorders are pathological conditions that directly affect voice production. Computer based diagnosis may play a major role in the early detection and in tracking and even development of efficient pathological speech diagnosis, based on a computerized acoustic evaluation. The health of the Voice is assessed by several acoustic parameters. The exactness of these parameters is often linked to algorithms used to estimate them for speech noise identification. That is why main effort of the scientists is to study acoustic parameters and to apply classification methods that achieve a high precision in discrimination. The primary aim of this paper is for a meta–analysis on voice disorder databases i.e. SVD, MEEI and AVPD and machine learning techniques applied on it. *Materials and Methods:* This field of study was systematically reviewed in compliance with PRISMA guidelines. A search was performed with a set of formulated keywords on three databases i.e. Science Direct, PubMed, and IEEE Xplore. A proper screening and analysis of articles were performed after which several articles were also excluded. *Results:* Forty-five studies that fulfills the eligibility criteria were included in this meta-analysis. After applying eligibility criteria on the peer reviewed and research article and studies that were published in authentic journals and conferences proceedings till June 2020 were chosen for further full-text screening. In general, only those articles that used voice recordings from SVD, MEEI and AVPD databases as a dataset is included in this meta-analysis. *Conclusion:* We discussed the strengths and weaknesses of SVD, MEEI and AVPD. After detailed analysis of the studies including the techniques used and outcome measurements, it was also concluded that Support Vector Machine (SVM) is the most common used algorithm for the detection of voice disorders. Other than was also noticed that researchers focus on supervised techniques for the clinical diagnosis of voice disorder rather than using unsupervised techniques. It was also concluded that more work needs to be on voice pathology detection using AVPD database.

1. Introduction

Speech problems are linked to negative effects on quality of life, significant indirect costs of speech-related work, short-term demands and projections of costs of primary health care approximately $5 billion a year all over the world [1]. Dysphony diagnosis may include medical therapy, surgery, and/or speech therapy. Speech therapy can either be used as the primary mode, as an alternative to or as a medium for treatment assistance. Voice therapy in patients with muscles tension dysphony and benign phono-traumatic vocal fold lesions, degeneration of the vocal folds associated with age, disorders of the neurons (incl. Parkinson's disease) and disorders of the voice associated with reflux was shown to be effective. To date, most studies of voice therapy have been carried out in university tertiary voice clinics, whereas further studies on use of speech therapy have been conducted by otolaryngologists who are subject to bias recall [1]. In what is perceived as a' normal voice' there is a huge variation. It is problematic to determine its essential properties because a continuum exists between a normal and a disordered voice. A normal voice is essentially in quality unnoticeable and allows sufficient communication and unnecessary effort or inconvenience. Hoarseness is a word that describes an abnormal, harsh, breathy, weak or strained voice quality. A voice problem or dysphony can be defined by structural or functional anomaly of the voice mechanism as any impaired, limited or restricted activity or participation in (world health organization) [2]. Vocal production of the voice may be specified by fundamental frequency, intensity, vibration and vocal intonation according to its vocal parameters. The perceptional correlates of frequency are known as pitch or subjective level sensations that are appropriate for age and sex and are known as loudness or subjective noise sensations that are suitable for the environment. [3]. A person's voice displays these features as gender, age, emotional state and cultural heritage [4]. This represents individual identity and makes it possible to differentiate between individuals. The voice represents different aspects of the individual's physical, social, cultural and psychological development at different stages of life infancy, puberty, adulthood and aging [5]. A good voice satisfies the professional and/or personal needs of the individual of full, and is held comfortable in a person's life. Expression quality may be affected by hormonal changes, asthma, disease, blood vessels, neurology and emotional disorders, operations or other general health-related factors [3]. There are however no universal criteria to determine the characteristics and limits of a normal voice and certain shifts in voice during a vocalization are anticipated and socially acceptable. But some developments cannot be as indicators of social or emotive expression, despite taking such changes into account. Such changes are then called dysphonia [4]. Voice disorders manifest in various ways, including the presence of sensory and auditory symptoms, deviations in vocal quality and functional and/or structural laryngeal changes that may involve behavioral and/or organic factors associated with their genesis and maintenance [5]. These disorders can have a negative impact on the patient's quality of life, compromising social, emotional, and work-related situations [6,7]. Patients with voice disorders may experience various symptoms, of which hoarseness, sore throat, vocal fatigue, and throat clearing are the most common. These symptoms may be associated with intense voice use, upper respiratory tract infections, stress, and smoking [8]. Because manifestation of a voice disorder is multidimensional, its assessment must include a variety of factors, including perceptual voice

assessment, visual laryngeal inspection, acoustic analysis, aerodynamic assessment, and vocal self-assessment [9]. Voice Pathology disorders can be detected using the classification tools for computer helped voice pathology. Language pathology recently focused on the techniques of machine learning. These tools can early diagnose and offer adequate treatment for voice pathologies. Clinical voice pathology is detected by several procedures, including acoustic analysis. Voice disorder services are available for the study of the auditory behavior of voices suffering from different forms of vocal disabilities in hospitals as much as in electronic voice disorder detection systems. The assessment of pain, such as dysphonia, is an essential factor of the medical evaluation and treatment of man's voice. In addition to larynx and vocal fold endoscopic testing, visual and acoustic measurement techniques are crucial components in the clinical evaluation of dysphonia. It consists of the calculation, in compliance with SIFEL Recommendations [10] Edicts and Phoniatrics, following the instructions of the Phoniatrics Committee of the European Society of Laryngology to identify certain modifications to the vocal tract, the relevant parameters obtained from the voice signal. It is, in contrast to other medical tests, a non - invasive clinical trial by direct observation of vocal folds, for example [11,12]. For medical diagnosis, the use of classifier systems slowly increases. The development of specialist networks and decision support (DSS) technologies for medical applications has led to the recent advancement in the field of artificial intelligence. Expert systems and various artificial detection intelligence methods had the ability to be good medical devices. Classification systems may contribute to the increase in precision, accuracy and reliability of diagnosis and the reduction of possible errors [13].

The first database that is used in this review is Saarbruecken Voice Database (SVD) [14]. A collection of voice recordings by over 2000 people. 1) Vocal registration [I a, u] produced at standard, high and low pitches. The truth was recorded in a recording session. 2) Vocal documentation of increasing pitch [I a, u]. 3) Recording of the phrase" Good morning, how do you like it?"(' How are you, good morning?'). The voice signal and the EGG signal were stored in individual files for the specified components. The database has text file includes all relevant information about the dataset. Those characteristics make it a good choice for experimenters to use. All recorded SVD voices were sampled with a resolution of 16-bit at 50 kHz. There are some recording sessions where not all vowels are included in each version, depending on the quality of their recording. The' Saarbruecken Voice Server' is available via this web interface. It contains multiple internet pages which are used to choose parameters for the database application, to play directly and records and pick the recording session files which are to be exported after chosen desired parameter from SVD database.

The second database that is used in this review is Massachusetts eye and ear infirmary (MEEI) [15]. Contains over 1,400 vocal tests of the long vowel / a/ and the first portion of the Rainbow passage, created by MEEI Voice and Speech Lab. It has been sold in two distinct surroundings by Kay Elemetrics [16]. The sampling frequency was 50 kHz, while the response frequency for normal samples was 25 kHz or 50 kHz, respectively. It is used in most voice pathology detection and classification experiments although the different conditions and sound levels used to capture normal and pathological voice have many drawbacks. In this collection, some tools, such as stroboscopy, auditory aerodynamics and physical neck and mouth tests, were used to assess speech disorders (this information was provided by Kay Elemetrics).

The third database that is used in this review is Arabic voice pathology database (AVPD) [15]. Samples of words and voices were recorded at various sessions in King Abdul Aziz University Hospital in Riyadh, Saudi Arabia, Communication & Swallowing Disorders Unit. In a sound treatment room, a standard recording protocol was used to collect voices of the patient by experienced phoneticists. The

database protocol has been developed to prevent specific MEEI data base deficiencies [17]. The AVPD provides records of long-standing vowels and voice folding disorders, coupled with the same records of regular speakers. After a laryngeal stroboscope has been clinically checked, pathological vocal folds have been identified. In the case of anatomy, the perceptive degree of voice disorders was calculated at a scale of 1–3, the most severe is 3. The gravity ranking of each sample was focused on the category of three medical experts. The texts are different: (1) three long-lasting vowels with initial details and offset details; (2) single Arabic and several common words; and (3) continuous speech. The chosen text has been specifically selected over all Arabic phonemes. Most speakers have reported three utterances of each vowel: /a/, /u/ and /i/. Just once single words and repetitive speaking were recorded to discourage patients from overloading them. For both normal and disease samples in AVPD, the test frequency is 50 kHz

This paper provides a meta-analysis of the relevant research articles that are directly targeting voice disorders and the databases use for the detection and the machine learning techniques used for the detection as explained in figure 1. This aim of this review is to investigates, summarizes, analyzes and discussions of a series of research articles regarding their details, finding and accuracy. Our research based on research papers from databases such as PubMed, IEEE Xplore and ScienceDirect, till June2020. In this paper, we primarily aim to assess the current efficacy of various methods of machine learning used to detect voice disorders and to explore the development, shortcomings and problems that have been made, as well as future research needs. To the best of our knowledge this is the first literature review that covers all three most popular databases i.e. SVD [14], MEEI [15] and AVPD [15] available for voice disorders. The important contributions of this paper are:

- Meta-analysis on the detection of voice disorder using SVD [14], MEEI [15] and AVPD [15] databases.
- Review outcomes and accuracy of 45 relevant articles.
- Identify the gap for research in this field.

The arrangement of this paper is organized as follows. Section 1 provides a short introduction of voice disorders and databases we have targeted. Section 2 provides the methodology used to conduct this review of the literature. The finding of this systematic assessment is mention in Section 3 of this paper. Section 4 deals with our main research concerns. This conclusion of this whole paper is provided in section 5 with restrictions, research gaps and recommendations for further investigations.

2. Methodology

2.1. Search strategy and database information and source

The population, intervention, comparison and outcome bases method (PICO) [18] was considered for this meta-analysis. The search strategy was set up according to PICO:

- P = (Population) = people with voice disorders
- I = (Intervention) = detection with data given in the form of voices. Here data extraction is done from SVD [14], MEEI [15] and AVPD [15].
- C = (Comparison) = different Machine learning algorithms
- O = (Outcome) = report accuracies and compare them.

A set of search strings was generated with the Boolean operator combining suitable synonyms and alternate terms: AND restricts and limits the quest and OR expands and extends the search [18]. With help of these Boolean operators the search term was formulated as: (voice disorder) AND (SVD/MEEI/AVPD) AND ("computer vision" OR "neural network" OR "artificial intelligence" OR "pattern recognition" OR "machine learning"). Peer-reviewed publications have been searched in 3 big databases: PubMed, IEEE Xplore and ScienceDirect. Search was restricted in ScienceDirect to review articles, research articles, conference abstracts, correspondences, data articles, discussions, case reports. All three databases have been searched till June 2020. The set of keywords were formulated that have been used to perform search in these databases. We searched these three databases three different time for each dataset we have target in our meta-analysis. The search results were in PubMed (n = 12), IEEE Xplore (n = 19) and ScienceDirect (n = 103). The total number of search results were (n = 134) when the initial searched was performed. Total included studies are 45 and this whole process has been explained in the figure 2 flowchart. The total number of each database used is for SVD (n = 20), MEEI (n = 31), AVPD (n = 6) and it has been represented in figure 1. With the help of pie chart. It can be seen from pie chart that MEEI is the most used database for voice pathologies detection.

Using the endnote web system, search results were stored and organized and a table of data extracted from every selected paper was created. For articles deemed to be potentially eligible, full texts were uploaded into the Endnote web (by Clarivate Analytics). The first search applied the search terms for each selected database and included the full document in both journals and conferences. Thousands of irrelevant findings have returned from this procedure, and therefore a decision is made to limit the search on the title and the type of content of the document. Further study is determined by reference to the sources of the related studies found. After collecting primary search studies, we scanned the titles and the abstract for the relevant studies. An ongoing investigation has been carried out with a complete text to assess the relevant studies.
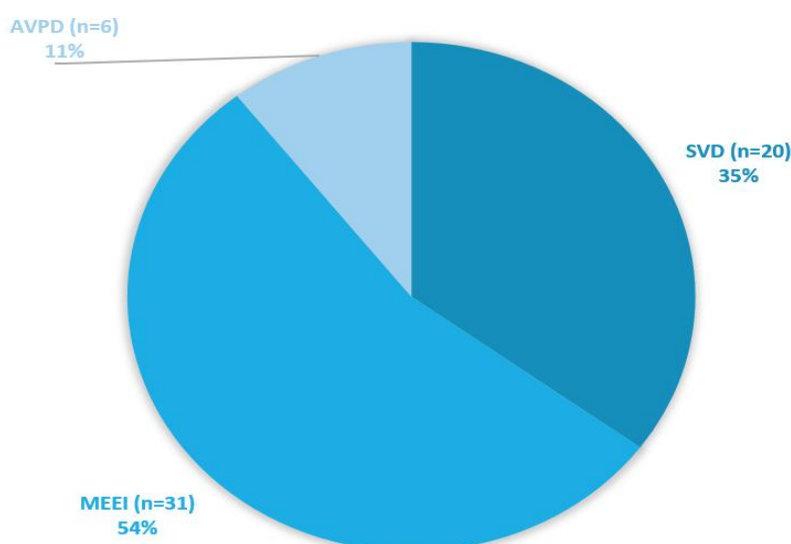

Figure 1. This pie chart represents no. of studies.

2.2. Eligibility criteria

This study focused on peer reviewed articles that used machine learning to recognize voice disorders in voice recordings as it's described in figure 2. In fact, we concentrated mostly on the related

research papers with respect to these criteria in order to understand the problem through machine learning or implementation. This only includes articles that solely used voice recordings from SVD [14], MEEI [15] and AVPD [15] database to detect voice disorder. The second criterion is to ensure that the selected research papers use approaches based on machine learning. The criteria eliminated any papers that do not include machine learning or an algorithm in which the disease is defined. This also excludes papers solely based on a qualitative examination and not analyzed on basis of accuracy and quantitative analyzes. The third criterion notes that the research papers chosen also include image detection software for disease. The criteria showed the accuracy of machine learning and its techniques applied in all selected article that are quantitatively reviewed published. In order to report irrelevant research papers, the inclusion and exclusion criteria were used. This examination paper outlines the inclusion and exclusion criteria used:

Inclusion Criteria:

- Research articles based on voice recordings as a data in order to predict the disorder.
- Studies that use any of the following: SVD [14], MEEI [15] and AVPD [15] database
- Research article consisting of machine learning techniques.
- Articles consists of voice filtering and segmentation techniques or an application or any software in order to detect the disease through voices.
- All articles are in the language of English.
- In either a journal or a conference proceeding published story is included.

Exclusion Criteria:

- Research article that do not include voice recordings as a data were excluded.
- Studies that didn't use SVD [14], MEEI [15] and AVPD [51] database.
- Research articles that do not use any machine learning.
- Articles that do not use voice filtering and segmentation are excluded.
- Research which have not been written in English.
- Research that were not included in any journal or conference proceedings.

3. Results

3.1. Identification of studies

From table 1 we can observe that all the selected and screened stories are in between 2002 to 2020. But most of the publications are from last five years which can be observed in figure 3 which proves that detection of voice disorders through machine learning techniques and to apply them in clinical setting is the area of interest for most of the researchers.

In table 1, it has been observed that SVM is the most used algorithm for the diagnosis of voice disorders in all three datasets. In our lives today the recognition of voice disorders plays an important role. Many of these disorders should therefore be treated until they progress to a critical condition at an early stage of incidence. SVMs have become a popular tool for discriminatory labeling. Speech synthesis is a promising field for recent SVM applications [64].
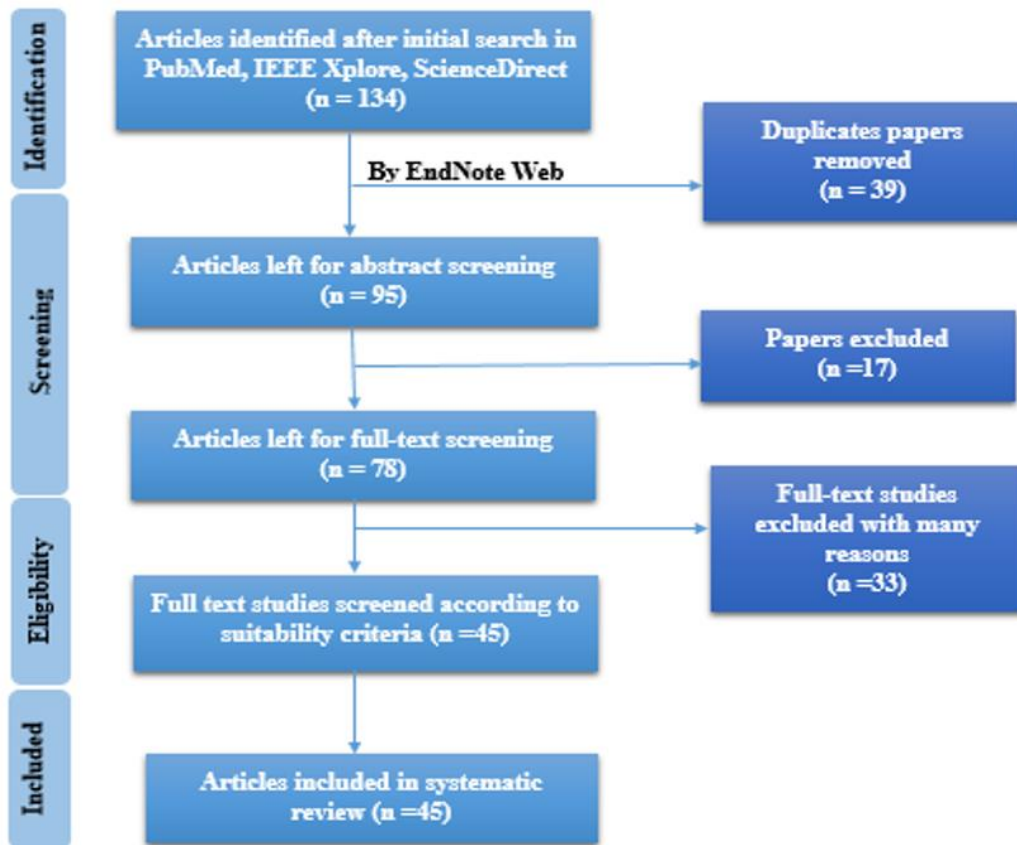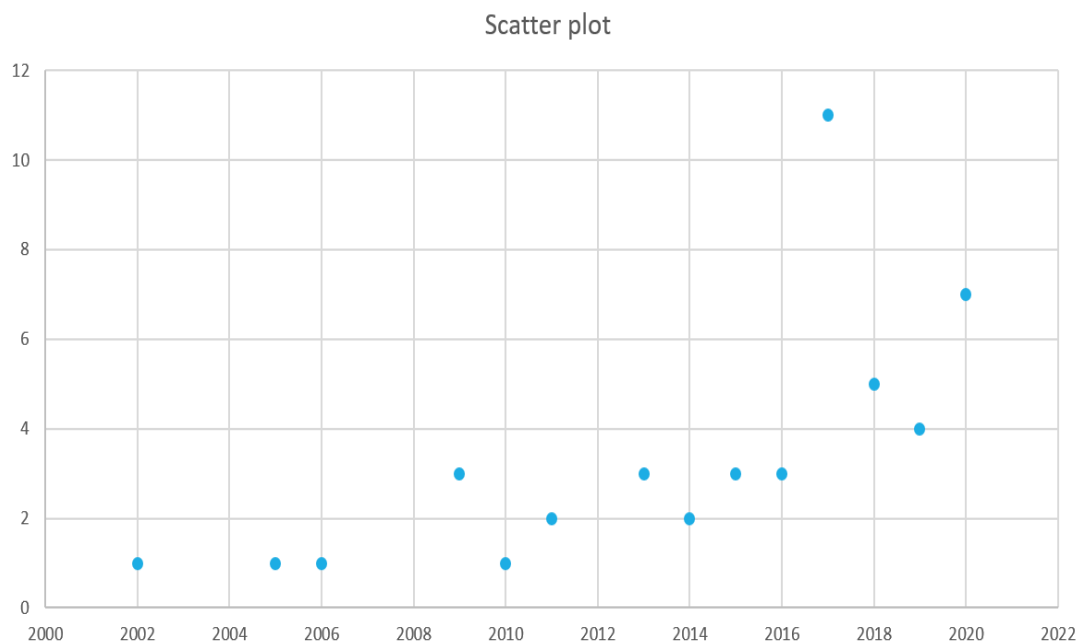
Figure 2. PRISMA Flowchart.

Figure 3. Scatter plot of publication year of databases.

Table 1. Summary of the 45 selected studies with data extraction.

No	Author/Year	ML technique/Classifier	Feature Selection/Filter	Overall Accuracy	Overall Sensitivity	Overall Specificity
			SVD Dataset			
1..	A. Al-Nasheri et al. /2017 [19]	SVM	peak, lag, entropy/eight band pass filter	99.53%	91.22%	94.27%
2.	Al-Nasheri et al./2017 [20]	SVM	MDVP parameters	99.68%	NA	NA
3.	Al-Nasheri et al./2017 [21]	SVM	Eight frequency bands using correlation functions	90.97%	NA	NA
4.	Zulfiqar Ali et al./2017 [22]	GMM	MFCCs	80.02%	71.0%–84.7%	70%–76.6%
5.	Fonseca et al./2020 [23]	SVM	SE, ZCRs, SH	95%	NA	NA
6.	Garcia et al./2019 [24]	Gaussian Mixture Regression	GBR scale	NA	NA	NA
7.	Guedes et al./2019 [25]	DLN (LSTM; CNN)	PCA	80; 78;66; 67;63; 66	80;78;66; 67;63; 66	80; 80;67; 67;69; 71
8.	Hammami et al./2020 [26]	SVM	HOS; DWT	99.3%; 93.1%	96.4%; 92.8%	99.4%; 93.3%
9.	Panek et al./2016 [27]	K-Means Clustering	PCA	100%	NA	NA
10.	Moon et al./2018 [28]	LR DT RF SVM DNN	HOS and DEO	82.77%;80.25%;84.87%;86.13%;87.4%	NA	NA
11.	Ezzine et al./2018 [29]	ANN SVM	glottal flow features	99.27%;98.43%	NA	NA
12.	Markaki et al./2009 [30]	RBF Kernal with SVM	Mutual Information b/w subjective voice quality and computed features	94.1%	NA	NA
13.	Markaki et al./2011 [31]	SVM	mutual information p/w voice classes (normophonic/dysphonic)	94.1%	NA	NA
14.	Miramont et al./2020 [32]	SVM	CPP, SDNPCV, NPCV, HNR,Noise Level, D2 (10, 25),Shimmer, and K2 (6, 25).	86.53%	NA	NA
15.	Muhammad etal. /2017 [33]	SVM	Glottal source excitation	93.2 ± 0.01	94.3	92.3
16.	Shia et al. /2017 [34]	FFNN	DWT	93.3%	NA	NA
17.	Kadiri et al./2020 [35]	SVM	Glottal source features and MFCC	76.19%	NA	NA
18.	Zhang et al./2020 [36]	DNN	Pitch extraction and line spectrum pair	NA	NA	NA

No	Author/Year	ML technique/Classifier	Feature Selection/Filter	Overall Accuracy	Overall Sensitivity	Overall Specificity
19.	Teixeira et al./2018 [37]	SVM	Jitter, shimmer and HNR, MFCC	71%	NA	NA
20.	Teixeira et al./2017 [38]	MLP−ANN	Jitter, shimmer and HNR	100%(female);90%(male)	NA	NA
	MEEI Dataset					
1.	A. A. Dibazar et al. /2002 [39]	HMM	me1 frequency filter	99.4% with E = 8%	NA	NA
2.	A. Al-Nasheri et al. /2017 [19]	SVM	peak, lag, entropy/eight band pass filter	99.54%	99.96%	99.96%
3.	Akbari et al./2014 [40]	MC-LDA; ML-NN	wavelet packet- based features	96.67%	NA	NA
4.	Al-Nasheri et al./2017 [20]	SVM	MDVP parameters	88.21%	88.90%	89.21%
5.	Al-Nasheri et al./2017 [21]	SVM	Eight frequency bands using correlation functions	99.80%	NA	NA
6.	Zulfiqar Ali et al./2016 [41]	GMM	Estimation of Auditory Spectrum and Cepstral Coefficients	99.56%	NA	NA
7.	Zulfiqar Ali et al./2017 [22]	GMM	MFCCs	94.6%	94.7%–99.1%	50.9%–94.5%
8.	Amami et al./2017 [42]	SVM	DBSCAN and MFCCs	98%	NA	NA
9.	Londono et al./2010 [43]	HMM	MFCCs	82.1472.2	81.1373.6	83.3373.4
10.	Arjmandi et al./2011 [44]	1)QD classfier 2)NM classifier 3)Parzen Classifier; KNN; SVM; ML NN	MDVP parameters	78.9%;87.20%;85.50% ; 88.86%;89.29%;88.7%	88%;70.9%;73.5%;78.30%;82.25%;83%	66%;97%;93.85%;96.17%;94.3%;85.1%
11.	Barreira et al./2020 [45]	Gaussian Naïve Bayes	HASS-KLD, H-KLD, MFCCs, Sample skewness	99.55%	100%	98%
12.	Francis et al./2016 [46]	ANN	MMTLS	96.48%	NA	NA
13.	Cordeiro et al./2017 [47]	SVM, GMM, DA	MFCC, LSF	98.7%	NA	NA
14.	Cordeiro et al./2018 [48]	SVM	RPPC	94.2%	NA	NA
15.	Fang et al. /2019 [49]	DNN SVM GMM	MFCCs	99.14 ± 1.9%;98.28 ± 2.3%;98.26 ±1.8%	NA	NA
16.	Muhammad et al. /2013 [14]	SVM	MPEG-7 low level audio feature	99.994% ±0.011	1	0.999

Continued on next page

No	Author/Year	ML technique/Classifier	Feature Selection/Filter	Overall Accuracy	Overall Sensitivity	Overall Specificity
17.	Muhammad et al. /2013 [50]	SVM	VTAI Feature Extraction	99.02% ± 0.01	99.8% ± 0.02	97.5% ± 0.04
18.	Ghasemzadeh et al. 2015 [51]	GA and LDA with SVM	Nonlinear features	98.4%	99.3 ± 1.2	94 ± 5.7
19.	Llorente et al./2009 [52]	MLP-NN	MFCCs	96 ± 1.3	0.99	0.82
20.	Hariharan et al./2013 [53]	LS-SVM; kNN PNN; CART	Wavelet packet transform based energy/entropy	92.24 ± 0.24;89.82 ± 0.28;89.54 ± 1.34;86.97 ± 0.20	93.02 ± 0.33;91.96 ± 0.56;90.62 ± 2.46;87.71 ± 0.28	91.49 ± 0.22;87.89 ± 0.43;88.59 ± 0.47;86.27 ± 0.42
21.	Ezzine et al./2018 [29]	ANN SVM	glottal flow features	93.6%,93.57%	NA	NA
22.	Mahmood /2019 [54]	Naïve Bayes ANN SVM RF	MFCC	72.70%,93.72%,99.78%,99.91%	NA	NA
23.	Mekyska et al./2015 [55]	SVM RF	spectra, inferior colliculus coefficients, bicepstrum, approximate entropy, empirical mode decomposition	99.9 ± 0.4 100.0 ± 0.0	99.8 ± 0.5 100.0 ± 0.0	99.9 ± 0.7 100.0 ± 0.0
24.	Miramont et al./2020 [32]	SVM	CPP, SDNPCV, NPCV, HNR, Noise Level, D2 (10, 25), Shimmer, and K2 (6, 25).	87.06%	NA	NA
25.	Muhammad et al. /2014 [56]	SVM	MPEG-7 feature	99.994%	1	0.999
26.	Muhammad et al. /2017 [33]	SVM	Glottal source excitation	99.4 ± 0.02	99.4%	98.9%
27.	Nayak et al./2005 [57]	ANN	DWT coefficients as a feature vector	80–85%	NA	NA
28.	Henriquez et al./2009 [58]	NN	first- and second- order Rényi entropies, correlation entropy, correlation dimension	99.69%	NA	NA
29.	Salehi et al./2015 [59]	SVM	Parametric wavelet by adaptation wavelet transform	98.30%	NA	NA
30	Lechon et al./2006 [60]	MFCC	NN	89.6 ± 2.49%	NA	NA
31.	Travieso et al./2017 [61]	HMM; Linear SVM; Kernal SVM	Nonlinear Dynamic Parameterization	93.55 ± 3.24;96.73 ± 3.42;99.87 ± 0.39	NA	NA
			AVPD Dataset			
1.	A. Al-Nasheri et al. /2017 [19]	SVM	peak, lag, entropy/eight band pass filter	96.02%	91.22%	94.27%

No	Author/Year	ML technique/Classifier	Feature Selection/Filter	Overall Accuracy	Overall Sensitivity	Overall Specificity
2.	Al-Nasheri et al./2017 [20]	SVM	MDVP parameters	72.53%	NA	NA
3.	Al-Nasheri et al./2017 [21]	SVM	Eight frequency bands using correlation functions	91.16%	NA	NA
4.	Zulfiqar Ali et al./2017 [22]	GMM	MFCCs	83.6%	67.9%–7.8.4%	75.9%–89.74%
5.	Mesallam et al./2017 [62]	SVM GMM VQ HMM	MFCC	93.6%,91.6%,90.3%, 88.9%	NA	NA
6.	Muhammad et al. /2017 [33]	SVM	Glottal source excitation	91.5 ± 0.09	92.2%	91.1%

ANN = Artificial Neural Network, **CART**= Classification and Regression Tree, **CNN** = Convolutional Neural Network, **DA** = Discriminant analysis, **DBSCAN** = Density Based Spatial Clustering of Applications with Noise. **DEO** = Differential Energy Operator, **DLN** = Deep Learning Network, **DNN** = Deep Neural Network, **DT** = Decision Tree, **DWT** = Discrete Wavelet Transform, **FFNN** = Feed Forward Neural Network, **GA** = Genetic Algorithm, **GMM** = Gaussian Mixture Model, **HASS-KLD** = Higher amplitude suppression spectrum Kullback–Leibler divergence, **H-KLD** = Histogram Kullback–Leibler Divergence, **HMM** = Hidden Markov Model, **HNR** = Harmonic to Noise Ratio, **HOS** = High Order Statistics features, **KNN** = K-Nearest Neighbor Classifier, **LDA** = Linear Discriminant analysis, **LR** = Logistic Regression, **LSF** = Line spectral frequencies, **LTSM** = Long Short Term Memory, **MC-LDA** = Multi-Class Linear Discriminant Analysis, **MDVP** = Multidimensional Voice Program parameters, **MFCCs** = Mel-frequency cepstral coefficients, **ML-NN** = Multilayer Neural Network, **MMTLS** = Modified Mellin Transform of Log Spectrum, **NA** = Not Available, **NM** = Nearest Mean Classifier, **NN** = Neural Network, **PCA**= Principal Component Analysis, **PNN** = Probabilistic Neural Network, **QD =** Quadratic Discriminant Classifier**, RF** = Random Forest, **RPPC** = Relative Power of the Periodic Component, **SE** = Signal Energy, **SH** = Signal Entropy, **SVM** = Support Vector Machine, **VQ** = Vector Quantizer, **VTAI** = Vocal Tract Area Irregularity, **ZCRs** = Zero-Crossing Rates.

Support Vector Machine (SVM) is an old classification approach and has shown great scientific interest, especially in the fields of machine classification, regression and learning. SVM with the known classes associated. This is defined as filtering or extraction of features. Even if no prediction of unknown samples is necessary, function selection and SVM classification have been used together. They may be used to define main sets that take part in the class differentiation process. The SVM maps the entrance space to a large area. The SVM could determine the border of areas belonging to both classes by calculating an optimal hyperplane separation. The hyperplane is chosen to maximize the distance between the nearest samples of workouts. Initially, SVM models have been defined to categorize linear classes. Because the area of characteristics is large, the function characteristics for finding the separation hyperplane cannot be used directly. The characteristic function is used to compute non-linear mapping using special non-linear functions known as the kernel. The Kernel has the advantage of working in the input area where the weighted sum of the kernel function evaluated by support vectors can be used to solve the classification problem. By using different kernel functions, the SVM algorithm can create a range of learning machines. SVM tends to have a far better accuracy and give promising results then artificial neural network [63]. SVM (support vector machines) have become a common tool for classification, regression or novelty recognition machine learning tasks. They demonstrate good performance in general terms on many real questions and the method is logically inspired. The design of the learner machine does not have to be sought through experimentation [66]. There are very few free parameters. While SVMs are extremely powerful classifiers utilizing non-linear kernels, there are some downsides to this: 1). To find the best model, various kernel configurations and model parameters must be tested; 2). Training can be very long, particularly if there are many features or examples in the data set; 3). It is difficult to understand their inner workings because the underlying models are based on complex mathematical structures and their findings are difficult to interpret. For eg, the selection of the features with all available data and the subsequent testing of classifier training yield a positive error estimate [65].
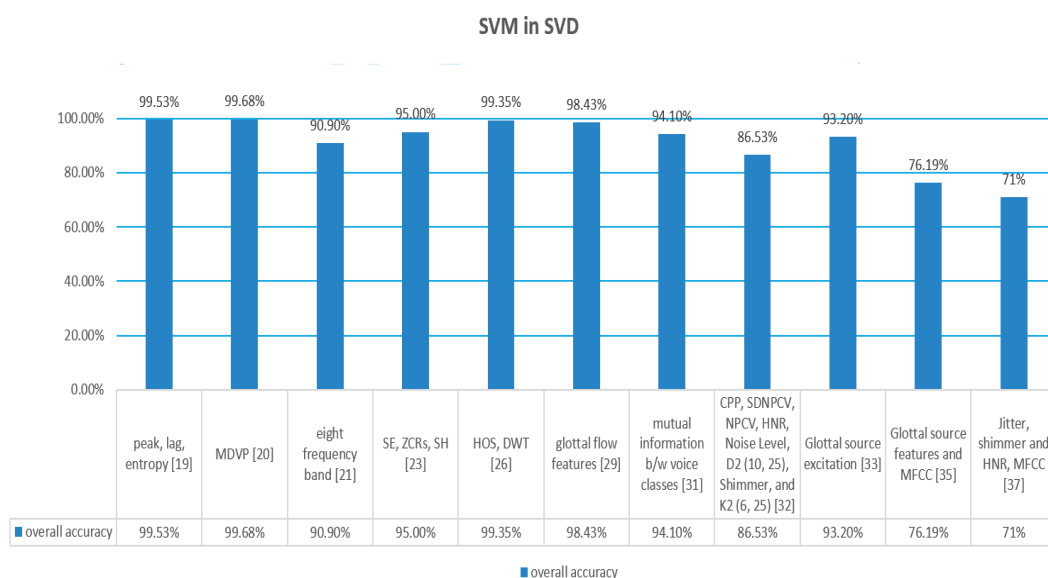

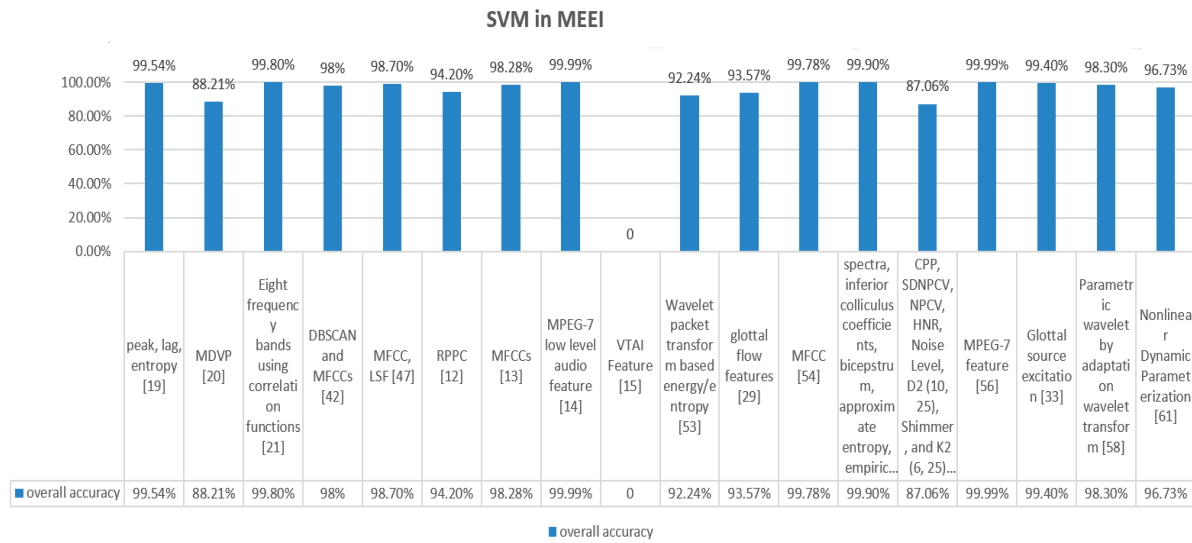**Figure 4.** Bar graph representing overall accuracy of SVM algorithm used in SVD with different features.

SVM in MEEI

	peak, lag, entropy [19]	MDVP [20]	Eight frequency bands using correlation functions [21]	DBSCAN and MFCCs [42]	MFCC, LSF [47]	RPPC [12]	MFCCs [13]	MPEG-7 low level audio feature [14]	VTAI Feature [15]	Wavelet packet transform based energy/entropy [53]	glottal flow features [29]	MFCC [54]	spectra, inferior colliculus coefficients, bicepstrum, approximate entropy, empiric...	CPP, SDNPCV, NPCV, HNR, Noise Level, D2 (10, 25), Shimmer, and K2 (6, 25)...	MPEG-7 feature [56]	Glottal source excitation [33]	Parametric wavelet by adaptation on wavelet transform [58]	Nonlinear Dynamic Parameterization [61]
overall accuracy	99.54%	88.21%	99.80%	98%	98.70%	94.20%	98.28%	99.99%	0	92.24%	93.57%	99.78%	99.90%	87.06%	99.99%	99.40%	98.30%	96.73%

■ overall accuracy

Figure 5. Bar graph representing overall accuracy of SVM algorithm used in MEEI with different features.

SVM in AVPD

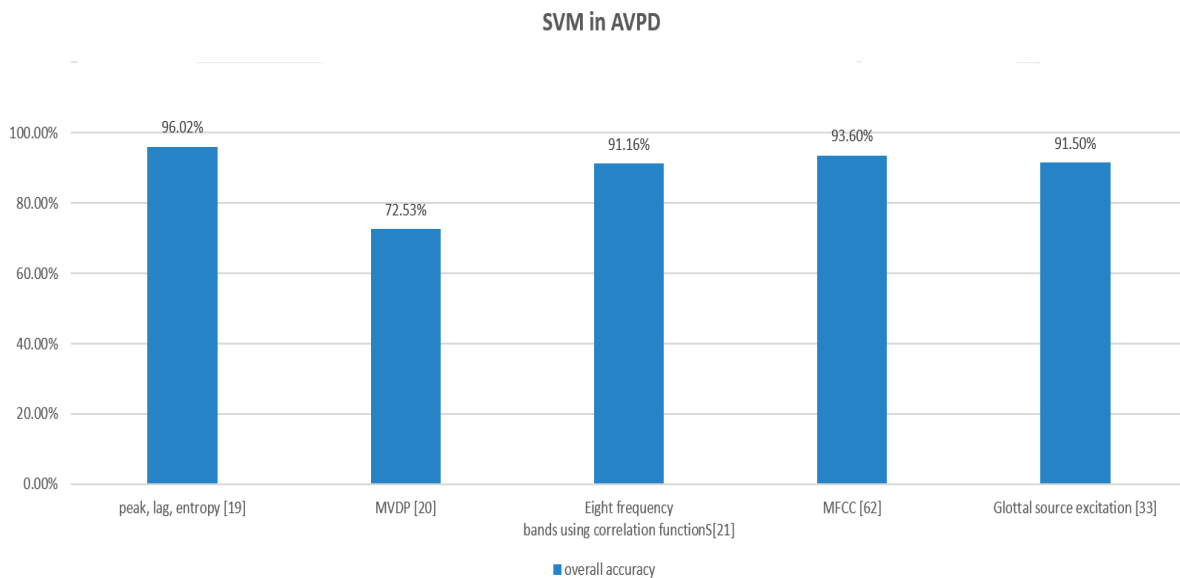| peak, lag, entropy [19]	MVDP [20]	Eight frequency bands using correlation functionS[21]	MFCC [62]	Glottal source excitation [33]
96.02%	72.53%	91.16%	93.60%	91.50%

■ overall accuracy

Figure 6. Bar graph representing overall accuracy of SVM algorithm used in AVPD with different features.

In figure 4, 5, and 6 a quantitative analysis has been carried out that shows that importance of SVM. SVM is the algorithm that has been widely used in the detection of voice disorders. For many years SVM and its application in the area of medical has been the topic of research for many researchers. SVM is the preference of scientist as a machine learning algorithm because of its best accuracy outcomes. In figure 4, 5, and 6 it has been observed that with variation in features different accuracies has been evaluated with SVM as a common algorithm in SVD [14], MEEI [15] and AVPD [15] database.
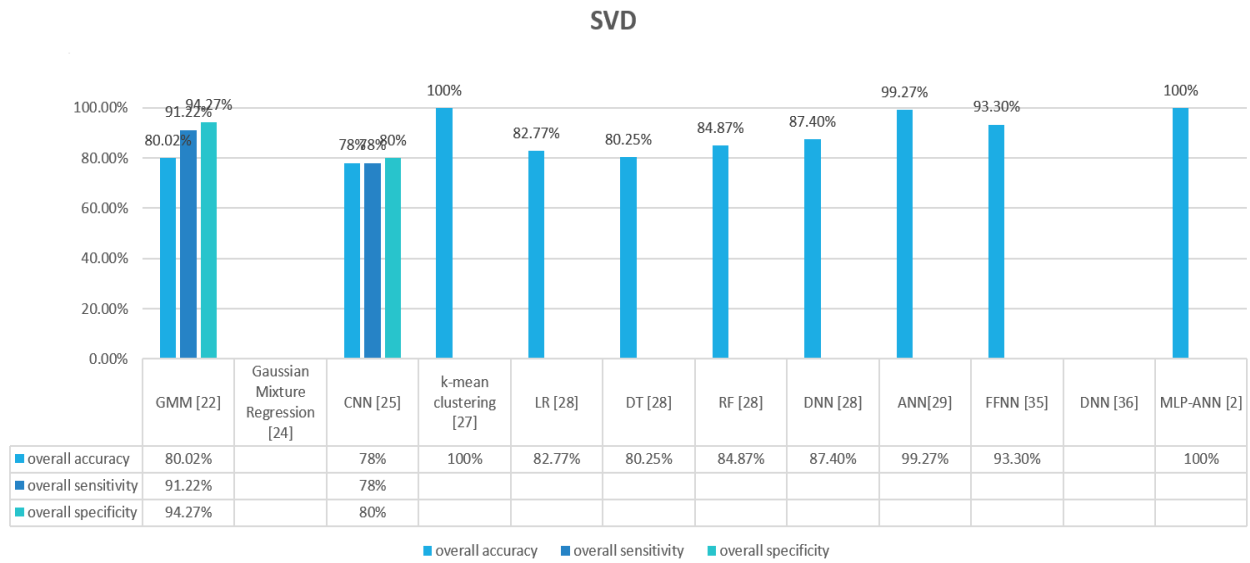
SVD

Figure 7. Bar graph representing overall accuracy of multiple algorithms used in SVD.
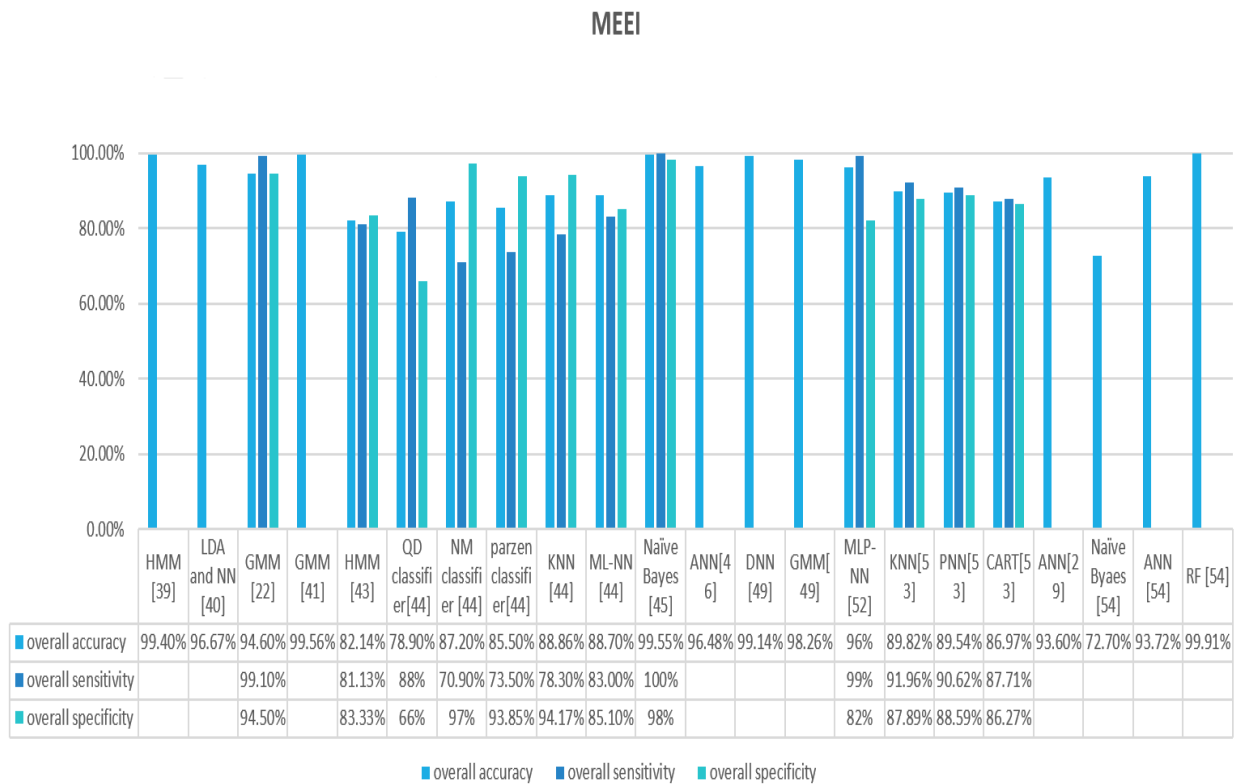
MEEI

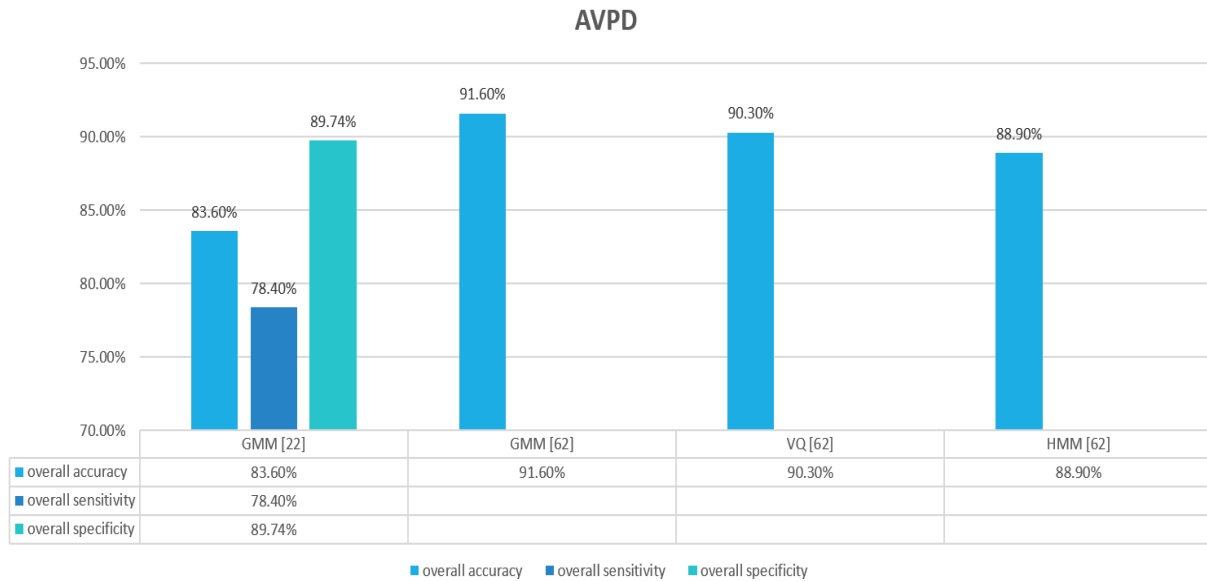**Figure 8.** Bar graph representing overall accuracy of multiple algorithms used in MEEI.

AVPD

	GMM [22]	GMM [62]	VQ [62]	HMM [62]
■ overall accuracy	83.60%	91.60%	90.30%	88.90%
■ overall sensitivity	78.40%			
■ overall specificity	89.74%			

■ overall accuracy ■ overall sensitivity ■ overall specificity

Figure 9. Bar graph representing overall accuracy of multiple algorithms used in AVPD.

In figure 7, 8, and 9 a quantitative analysis has been carried out between other algorithms in all selected databases. It has been observed that other than SVM, there are some algorithms that are resulted in good accuracies. For example, in graph 5 of SVD, Zulfiqar Ali et al. [22], GMM is used and the resulted accuracy is 80.02% with sensitivity 91.22% and specificity 94.27%. A Gaussian mixture model (GMM), as a weighted sum of Gaussian elements, is a parametric probability density function. GMMs are commonly used as a parametric model to distribute the probability in continuous measurements or characteristics in a biometry system, such as spectral related vocal-tract characteristics in a speech recognition system. The GMM parameters can be estimated from training data based on a well-qualified pre-model iterative EM or Maximum Posteriori (MAP) estimation [67]. In Moon et al. [28], Random Forest algorithm is used to detect voice disorders and the resulted outcome is 84.87% accuracy however overall sensitivity and specificity were not reported. RF is a series or community of classification trees and regression trees [68] which is trained in datasets of the same scale as the training set, called bootstraps. Once a tree is developed, bootstraps are used as test set which do not contain any specific record of the original (out - of-bag (OOB)) samples. The OOB estimate of the generalization error is the error rate of classification in all test sets. In 1996 [69] Breiman found that an OOB mistake is correct with a test set of the same size as that for the bagged classifiers. It removes the need for a different test set with the OOB calculation. In SVD, the highest reported accuracy of is 99% [20]. After SVM, GMM [22,24] and RT [29], convolutional neural network used in the detection of voice disorder and resulted in good outcome. A class that is influential in various computer vision tasks, Convolutional neural network (CNN) is attracting interest through a range of domains, including radiology. CNN is designed to learn spatial hierarchies through numerous building blocks, including cooling layers, bonding layers and fully connected layers, automatic and adaptive context propagation. [70]. CNN is a deep learning method that is commonly used for solving difficult problems. CNN is a deep learning solution. This overcomes the limitations of traditional machines [71]. In [25] CNN is used and the reported accuracy is 78%.

In figure 8 of MEEI, Naïve Bayes [54] has the lowest reported accuracy which is 72.70%. Other

than Naïve Bayes, algorithms like HMM [39,43], LDA [40], GMM [22,41,49], RF [54], PNN [53], KNN [53],

ANN [49,29] all have accuracies ranging in between 90% to 100%, which is again considered as the good reported outcome in terms of accuracy.

SVD [14], MEEI [15] and AVPD [15] databases are the center focus of this meta-analysis. Table 2 contain the basic differences in between all three databases which include their language, location, sampling frequency and the text that has been recorded.

Table 2. Comparative chart of SVD [61], MEEI [1] and AVPD [1] database.

Comparative Characteristic	SVD	MEEI	AVPD
language	German	English	Arabic
Location	Saarland University, Germany	Massachusetts Eye & Ear Infirmary (MEEI) voice and speech laboratory, USA	King Abdul-Aziz University Hospital, Saudi Arabia
Sampling frequency	50 KHz	10 KHz 25 KHz 50 KHz	48 KHz
Text	Vowel /a/ -Vowel /i/ -Vowel /u/ -sentence	-Vowel /a/ -Rainbow passage	-Vowel /a/ -Vowel /i/ -Vowel /u/ -Al-Fateha -Arabic digits -Common words

In pathology evaluation, perceptual severity has a major role to play, which either in SVD or MEEI repositories is not accessible. A confusion matrix provides information on honestly and incorrect categorized topics in an automated disturbance detection system. The cause for misclassification can be calculated by the perceptual severity of this structure. Automatic systems can at times not differentiate between typical abnormal subjects and relatively severe ones. This is why the perceptive severity in the AVPD is also taken into account in grades 1–3, in which 3 is a highly severe speech disorder. In comparison the typical AVPD participants are reported in the same state as those used for the pathological subjects following the clinical assessment [76]. A clinical examination of standard MEEI topics is not conducted although the history of the speech problem is incomplete [72]. No such information is provided in the SVD database. In AVPD, according to the MEEI database, all normal and pathological specimens are recorded at a single AVPD sampling frequency. Deliyski et al. concluded that the precision and the efficiency of the acoustic analysis is affected by the frequency of the sampling [73]. However, there is a vowel in the MEEI database and three vowels are registered in the AVPD. While three vows are also recorded in the SVD, they are only reported once. In the AVPD, three vowels are repeatedly reported, as some studies have suggested to model the intraspeaker variability for more than one single sample of the same vowel [74,75]. The total length of the reported study, that is 60 seconds, is another important feature of the AVPD. By regular as well as disordered individuals any text reported in an AVPD is of the same duration. Between normal and pathologic

topics, the recording times in the MEEI database vary. In comparison, the connected language (sentence) duration in the SVD database is only 2 seconds, which is not enough to build an automatic speech detection system. In addition, the SVD database cannot be used for a text-independent system. The AVPD is 18 seconds long on average and comprises seven sentences. The length of Al-Fateha speech is 18 seconds and it is segmented into two components to develop text-independent structures [76].

4. Discussion

After detailed quantitative analysis it has been noticed that only one unsupervised technique is used and that is only in SVD in Panek et al. /2016 [27] and its resulted accuracy is up to 99% although resulted sensitivity and specificity is missing. Other than no researcher has used any unsupervised technique for voice pathology detection. The validation of PCA by k-mean clustering and cross validation loses 10% signal (the variance of 90%) from the initial vector of the feature and produces worse results than the analysis by the original 28 vectors of functionality. In comparison with the results for women, the analysis based on kPCA included all the pitches analyzed showed the most accurate evidence of patient's health and condition. The analogous analysis of male recordings showed 100 % accuracy for 28 feature vectors and for the relevant number of key components for each pitch and kPCA result for each vowel. The k-means algorithm provides perfect separation of data for male recordings, which is the opposite of the female analysis using 28 parameters and PCA. This question was coped to and 99% of the classification accuracy from the kPCA analytics, which are non-linear data transformation. This indicates that the isolation of data in linear fashion was not adequate. In addition, k-means algorithm is presented as artifacts allocated by distance to the closest cluster. [27], though it is been suggested that researchers should focus more unsupervised techniques and evaluate these databases.

Tissue diseases, systemic changes, mechanical stress, surface discomfort, change in tissue, changes in neurology and muscle, and other factors [53] can cause Voice disease. The agility, strength and form of Vocal folds, resulting in abnormal noise and reduced acoustic tone, was affected by the vocal pathology. Subjective and objective evaluations of vocal problems have been approached until now [78]. The first group (subjective assessment) is the auditory and visual analysis of vocal folds in a hospital [77]. The first is a subjective assessment. The second category (target evaluation) is focused on automatic computer-based processing of acoustic signals to measure and identify the underlying vocal pathology, which may not even be detected by a human [62]. Therefore, this type of assessment is inherently non-subjective. Within reality, voices can now easily be captured and stored globally via cloud technologies using many intelligent devices. Many libraries have been commonly used by researchers for the objective assessment of speech pathology. The Massachusetts Eye and Ear Infirmary (MEEI) [15], the Saarbrücken Voice Database (SVD) [14], and the Arabic Voice Pathology Database (AVPD) [15]. In the repositories there are also some pitfalls. For example, certain bases are highly uniformly distributed within stable and unhealthy groups, and datasets provide troubling differences in the number of samples per type of pathology (e.g. there are fewer than 3 as more pathologies in the database). Some repositories do not have details on the severity of disease or on pathology symptoms during phonation, so some of the samples may seem safe, despite being called pathology and vice versa. Not to mention that more than 1 type of pathology is used to label documents and it is particularly challenging to incorporate or delete samples in different language [77].

Talking about the limitation of this systematic review, we cannot deny the fact of lower number

of included publications. Secondly those articles were selected which were published in English language, which can restrict the portrayal of work from non-English speaking countries and limit the generalizability of the results. Thirdly, there's a big possibility that search strategy for this review may have missed some relevant studies, since the studies which were published in conference proceedings were avoided mostly.

5. Conclusion

We discussed the strengths and weaknesses of SVD, MEEI and AVPD. After detailed analysis of the studies including the techniques used and outcome measurements, it was also concluded that Support Vector Machine (SVM) is the most common used algorithm for the detection of voice disorders. The amount of work done in this field concluded that clinical diagnosis voice disorders through machine learning algorithms have been the area of interest for most researchers. Other than was also noticed that researchers focus on supervised techniques for the clinical diagnosis of voice disorder rather than using unsupervised techniques. The identified gap that researchers should also focus more on unsupervised techniques in future so the analysis can be made based on their results that which provides the best outcomes and results. The second identified gap is that more work needs to be done on the AVPD database to evaluate its data with more feature extraction.

Conflict of interest

The authors have no conflict of interest in the conducted study.

References

1. S. Misono, S. Marmor, N. Roy, T. Mau, S. Cohen, Multi-institutional study of voice disorders and voice therapy referral, *Otolaryngol. Head Neck Surgery*, **155** (2016), 33–41.
2. P. Bradley, Voice disorders: Classification, *Otolaryngol. Head Neck Surgery*, (2010), 555–562.
3. M. Behlau, M. L. S. Dragone, L. Nagano, The voice that teaches: The teacher and oral communication in the classroom, 2004.
4. A. E. Aronson, Clinical voice disorders, 3 ed., INC. New York: Thieme Medical Publishers, 1990, p. 3–11.
5. J. R. Spiegel, R. T. Sataloff, K. A. Emerich, The young adult voice, *J. Voice,* **11** (1997), 138–143.
6. L. O. Ramig, K. Verdolini, Treatment efficacy: Voice disorders, *J. Speech Lang. Hear. Res.,* **41** (1998), 101–106.
7. J. Baker, The role of psychogenic and psychosocial factors in the development of functional voice disorders, *J. Speech Lang. Pathol.*, **10** (2008), 210-230.
8. S. T. Kasama, A. G. Brasolotto, Vocal perception and life quality, *Pro. Fono.,* **9** (2007), 19–28.
9. L. P. Ferreira, J. G. Santos, M. F. B. Lima, Vocal sympton and its probable cause: Data colleting in a population, *Rev. CEFAC,* **11** (2009), 110–118.
10. P. H. Dejonckere, P. Bradley, P. Clemente, G. Cornut G, L. C. Buchman, G. Friedrich, 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.,* **258** (2001), 77–82.

11. U. Cesari, G. De Pietro, E. Marciano, C. Niri, G. Sannino, L. Verde, Voice disorder detection via an m-Health system: Design and results of a clinical study to evaluate Vox4Health, *BioMed. Res. Int.,* **2018** (2018), 1–19.

12. L. Verde, G. De Pietro, G. Sannino, Voice disorder identification by using machine learning techniques, *IEEE Access*, **6** (2018), 16246–16255.

13. A. G. David, J. B. Magnus, Diagnosing parkinson by using artificial neural networks and support vector machines, *Global J. Comput. Sci. Technol.,* (2009), 63–71.

14. Saarbruecken Voice Database—Handbook, Stimmdatenbank.coli.uni-saarland.de. [Online]. Available: http://www.stimmdatenbank.coli.uni-saarland.de/help_en.php4.

15. M. OpenCourseWare, Lab Database | Laboratory on the Physiology, Acoustics, and Perception of Speech | Electrical Engineering and Computer Science | MIT OpenCourseWare, Ocw.mit.edu. [Online]. Available: https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-542j-laboratory-on-the-physiology-acoustics-and-perception-of-speech-fall-2005/lab-database/

16. K. Daoudi, B. Bertrac, On classification between normal and pathological voices using the MEEI-KayPENTAX database: Issues and consequences, INTERSPEECH-2014, Sep 2014, Singapour, Singapore. ffhal-01010857

17. N. Sáenz-Lechón, J. I. Godino-Llorente, V. Osma-Ruiz, P. Gómez-Vilda, Methodological issues in the development of automatic systems for voice pathology detection, *Biomed. Signal Process. Control*, **1** (2006), 120–128.

18. A. Liberati, D. G. Altman, J. Tetzlaff, C. Mulrow, P. C. Gøtzsche, J. P. A. Ioannidis, et al., The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration, *BMJ*, **339** (2009).

19. A. Al-Nasheri, G. Muhammad, M. Alsulaiman, Z. Ali, K. H. Malki, T. A. Mesallam, et al., Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions, *IEEE Access*, **6**, 6961–6974.

20. A. Al-Nasheri, G. Muhammad, M. Alsulaiman, Z. Ali, T. A. Mesallam, M. Farahat, et al., An investigation of multidimensional voice program parameters in three different databases for voice pathology detection and classification, *J. Voice*, **31** (2017), 113.e9–e18.

21. A. Al-Nasheri, G. Muhammad, M. Alsulaiman, Z. Ali, Investigation of voice pathology detection and classification on different frequency regions using correlation functions, *J. Voice*, **31** (2017), 3–15.

22. Z. Ali, M. Alsulaiman, G. Muhammad, I. Elamvazuthi, A. Al-Nasheri, T. A. Mesallam, K. H. Malki, et al., Intra- and inter-database study for Arabic, English, and German databases: Do conventional speech features detect voice pathology?, *J. Voice*, **31** (2017), 386.e1–e8.

23. E. S. Fonseca, R. C. Guido, S. B. Junior, H. Dezani, R. R. Gati, D. C. Mosconi Pereira, Acoustic investigation of speech pathologies based on the discriminative paraconsistent machine (DPM), *Biomed. Signal Process. Control*, **55** (2020).

24. J. A. Gómez-García, L. Moro-Velázquez, J. Mendes-Laureano, G. Castellanos-Dominguez, J. I. Godino-Llorente, Emulating the perceptual capabilities of a human evaluator to map the GRB scale for the assessment of voice disorders, *Eng. Appl. Artific. Intell.,* **82** (2019), 236–-251.

25. V. Guedes, F. Teixeira, A. Oliveira, J. Fernandes, L. Silva, A. Junior, et al., Transfer Learning with AudioSet to Voice Pathologies Identification in Continuous Speech, *Proced. Comput. Sci.,* **164** (2019), 662–669.

26. I. Hammami, L. Salhi, S. Labidi, Voice pathologies classification and detection using EMD- DWT analysis based on higher order statistic features, *IRBM*, **41** (2020), 161–171.

27. D. Hemmerling, A. Skalski, J. Gajda, Voice data mining for laryngeal pathology assessment, *Comput. Biol. Med.,* **69** (2016), 270–276.

28. J. Moon, S. Kim, An approach on a combination of higher-order statistics and higher-order differential energy operator for detecting pathological voice with machine learning, 2018 International Conference on Information and Communication Technology Convergence (ICTC), 17–19 Oct. 2018, pp. 46–51.

29. K. Ezzine, M. Frikha, Investigation of glottal flow parameters for voice pathology detection on SVD and MEEI databases, 2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 21–24 March 2018, pp. 1–6.

30. M. Markaki, Y. Stylianou, Using modulation spectra for voice pathology detection and classification, 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 3–6 Sept. 2009, pp. 2514–2517.

31. M. Markaki, Y. Stylianou, Voice pathology detection and discrimination based on modulation spectral features, *IEEE Transact. Aud. Speech Langu. Process*, **19** (2011), 1938–1948.

32. J. M. Miramont, J. F. Restrepo, J. Codino, C. Jackson-Menaldi, G. Schlotthauer, Voice signal typing using a pattern recognition approach, *J. Voice*, 2020.

33. G. Muhammad, M. Alsulaiman, Z. Ali, T. A. Mesallam, M. Farahat, K. H. Malki, et al., Voice pathology detection using interlaced derivative pattern on glottal source excitation, *Biomed. Signal Process. Control*, **31** (2017),156–164.

34. S. E. Shia, T. Jayasree, 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), 23–25 March 2017, pp. 1–6.

35. S. R. Kadiri, P. Alku, Analysis and detection of pathological voice using glottal source features, *IEEE J. Select. Topics Signal Process.*, **14** (2020), 367–379.

36. T. Zhang, Y. Shao, Y. Wu, Z. Pang, G. Liu, Multiple vowels repair based on pitch extraction and line spectrum pair feature for voice disorder, *IEEE J. Biomed. Health Inform.,* **24** (2020), 1940–1951.

37. F. Teixeira, J. Fernandes, V. Guedes, A. Junior, J. P. Teixeira, Classification of control/pathologic subjects with support vector machines, *Proced. Comput. Sci.,* **138** (2018), 272–279.

38. J. P. Teixeira, P. O. Fernandes, N. Alves, Vocal acoustic analysis—classification of dysphonic voices with artificial neural networks, *Proced. Comput. Sci.*, **121** (2017), 19–26.

39. G. Muhammad, M. Melhem, Pathological voice detection and binary classification using MPEG-7 audio features, *Biomed. Signal Process. Control*, **11** (2014), 1–9.

40. J. Nayak, P. S. Bhat, R. Acharya, U. V. Aithal, Classification and analysis of speech abnormalities, *ITBM-RBM*, **26** (2005), 319–327.

41. Z. Ali, I. Elamvazuthi, M. Alsulaiman, G. Muhammad, Automatic voice pathology detection with running speech by using estimation of auditory spectrum and cepstral coefficients based on the all-pole model, *J. Voice*, **30** (2016), 757.e7–e19.

42. R. Amami, A. Smiti, An incremental method combining density clustering and support vector machines for voice pathology detection, *Comput. Electr. Eng.,* **57** (2017), 257–265.

43. J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, G. Castellanos-Domínguez, An improved method for voice pathology detection by means of a HMM-based feature space transformation, *Patt. Recogn.,* **43** (2010), 3100–3112.

44. M. K. Arjmandi, M. Pooyan, M. Mikaili, M. Vali, A. Moqarehzadeh, Identification of voice disorders using long-time features and support vector machine with different feature reduction

methods, *J. Voice*, **25** (2011), e275–e289.

45. R. R. A. Barreira, L. L. Ling, Kullback–leibler divergence and sample skewness for pathological voice quality assessment, *Biomed. Signal Process. Control*, 57 (2020), 101697.

46. C. R. Francis, V. V. Nair, S. Radhika, A scale invariant technique for detection of voice disorders using Modified Mellin Transform, 2016 International Conference on Emerging Technological Trends (ICETT), 21–22 Oct. 2016, pp. 1–-6.

47. H. Cordeiro, J. Fonseca, I. Guimarães, C. Meneses, Hierarchical classification and system combination for automatically identifying physiological and neuromuscular laryngeal pathologies, *J. Voice*, **31** (2017), 384.

48. H. T. Cordeiro, C. M. Ribeiro, Spectral envelope first peak and periodic component in pathological voices: A spectral analysis, *Proced. Comput. Sci.,* **138** (2018), 64–71.

49. S. H. Fang, Y. Tsao, M. J. Hsiao, J. Y. Chen, Y. H. Lai, F. C. Lin, et al., Detection of pathological voice using cepstrum vectors: A deep learning approach, *J. Voice*, **33** (2019), 634–641.

50. G. Muhammad, Voice pathology detection using vocal tract area, 2013 European Modelling Symposium, 20–22 Nov. 2013, pp. 164–168.

51. H. Ghasemzadeh, M. Tajik Khass, M. Khalil Arjmandi, M. Pooyan, Detection of vocal disorders based on phase space parameters and Lyapunov spectrum, *Biomed. Signal Process. Control*, **22** (2015), 135–145.

52. J. I. Godino-Llorente, R. Fraile, N. Sáenz-Lechón, V. Osma-Ruiz, P. Gómez-Vilda, Automatic detection of voice impairments from text-dependent running speech, *Biomed. Signal Process. Control*, **4** (2009), 176–182.

53. M. Hariharan, K. Polat, R. Sindhu, S. Yaacob, A hybrid expert system approach for telemonitoring of vocal fold pathology, *Appl. Soft Comput.,* **13** (2013), 4148–4161.

54. A. Mahmood, A solution to the security authentication problem in smart houses based on speech, *Proced. Comput. Sci.,* **155** (2019), 606–611.

55. J. Mekyska, E. Janousova, P. Gomez-Vilda, Z. Smekal, I. Rektorova, I. Eliasova, et al., Robust and complex approach of pathological speech signal analysis, *Neurocomputing,* **167** (2015), 94–111.

56. G. Muhammad, M. Melhem, Pathological voice detection and binary classification using MPEG-7 audio features, *Biomed. Signal Process. Control*, **11** (2014), 1–9.

57. J. Nayak, P. S. Bhat, R. Acharya, U. V. Aithal, Classification and analysis of speech abnormalities, *ITBM-RBM*, **26** (2005), 319–327.

58. P. Henriquez, J. B. Alonso, M. A. Ferrer, C. M. Travieso, J. I. Godino-Llorente, F. Diaz-de- Maria, Characterization of healthy and pathological voice through measures based on nonlinear dynamics, *IEEE Transact. Audio Speech Lang. Process.*, **17** (2009), 1186–1195.

59. P. Salehi, Using patient's speech signal for vocal ford disorders detection based on lifting scheme, in 2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI), 5–6 Nov. 2015, pp. 561–568.

60. N. Sáenz-Lechón, J. I. Godino-Llorente, V. Osma-Ruiz, P. Gómez-Vilda, Methodological issues in the development of automatic systems for voice pathology detection, *Biomed. Signal Process. Control*, **1** (2006), 120–128.

61. C. M. Travieso, J. B. Alonso, J. R. Orozco-Arroyave, J. F. Vargas-Bonilla, E. Nöth, A. G. Ravelo-García, Detection of different voice diseases based on the nonlinear characterization of speech signals, *Expert Systems Appl.*, **82** (2017), 184–195.

62. T. A. Mesallam, F. Mohamed, K. H. Malki, A. Mansour, A. Zulfiqar, A. N. Ahmed, et al.,

Development of the arabic voice pathology database and its evaluation by using speech features and machine learning algorithms, *J. Healthc. Eng.*, (2017), 1–13.

63. K. Uma Rani, Mallikarjun S Holi, A comparative study of neural networks and support vector machines for neurological disordered voice classification, *Inter. J. Eng. Res. Techol.*, **3** (2014).

64. J. Godino-Llorente, P. Gómez-Vilda, N. Sáenz-Lechón, M. Blanco-Velasco, F. Cruz-Roldán, M. Ferrer-Ballester, Support vector machines applied to the detection of voice disorders, *Nonlin. Analy. Algor. Speech Process.*, (2006), 219–230.

65. S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, W. Xu, Applications of support vector machine (svm) learning in cancer genomics, *Cancer Genom. Proteom.,* **15** (2018).

66. S. Yue, P. Li, P. Hao, SVM classification: Its contents and challenges, *Appl. Math. J. Chinese Univer.,* **18** (2003), 332–342.

67. D. Reynolds, Gaussian Mixture Models, In: S. Z. Li, A. Jain (eds), Encyclopedia of Biometrics, Springer, Boston, MA, 2009.

68. L. Breiman, J. Friedman, C. J. Stone, R. A. Olshen, Classification and regression trees, Boca Raton, FL: CRC press, 1984.

69. L. Breiman, Bagging predictors, *Mach. Learn.,* 24 (1996), 123–140.

70. S. Indolia, A. Goswami, S. Mishra, P. Asopa, Conceptual understanding of convolutional neural network- A deep learning approach, *Proced. Computer Sci.*, **132** (2018), 679–688.

71. R. Yamashita, M. Nishio, R. Do, K. Togashi, Convolutional neural networks: An overview and application in radiology, *Insights Imag.,* **9** (2018), 611–629.

72. V. Parsa, D. G. Jamieson, Identification of pathological voices using glottal noise measures, *J. Speech Langu. Hear. Res.,* **43** (2000), 469–485.

73. D. D. Deliyski, H. S. Shaw, M. K. Evans, Influence of sampling rate on accuracy and reliability of acoustic voice analysis, *Logoped. Phoniatr. Vocol.*, **30** (2005), 55–62.

74. Y. Horii, Jitter and shimmer in sustained vocal fry phonation, *Folia Phoniatr.*, **37** (1985), 81–86.

75. J. L. Fitch, Consistency of fundamental frequency and perturbation in repeated phonations of sustained vowels, reading, and connected speech, *J. Speech Hear. Disord.*, **55** (1990), 360–363.

76. T. Mesallam, M. Farahat, K. Malki, M. Alsulaiman, Z. Ali, A. Al-nasheri, et al., Development of the arabic voice pathology database and its evaluation by using speech features and machine learning algorithms, *J. Healthc. Eng.,* 2017, 1–13.

77. P. Harar, Z. Galaz, J. Alonso-Hernandez, J. Mekyska, R. Burget, Z. Smekal, Towards robust voice pathology detection, *Neural Comput. Appl.,* 2018.

78. D. D. Mehta, R. E. Hillman, Voice assessment: Updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods, *Curr. Opin. Otolaryngol. Head Neck Surg.,* **16** (2008), 211.