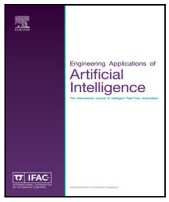
Engineering Applications of Artificial Intelligence 133 (2024) 108047

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Research paper

Voice disorder detection using machine learning algorithms: An application in speech and language pathology

Mujeeb Ur Rehman [a,*], Arslan Shafique [b], Qurat-Ul-Ain Azhar [c], Sajjad Shaukat Jamal [d], Youcef Gheraibia [e], Aminu Bello Usman [f]

[a] School of Computer Science and Informatics, Institute of Artificial Intelligence, De Montfort University, UK
[b] School of Biomedical Engineering, University of Glasgow, UK
[c] Department of Speech and Language Pathology, Riphah International University, Pakistan
[d] Department of Mathematics, College of Science, King Khalid University, Abha, 61413, Saudi Arabia
[e] School of Computer Science and Informatics, De Montfort University, UK
[f] School of Computer Science, University of Sunderland, UK

ARTICLE INFO

Keywords:
Voice signals
Intelligent systems
Support vector machine
Speech features

ABSTRACT

The healthcare industry is currently seeing a significant rise in the use of mobile devices. These devices not only provide ways for communication and sharing of multimedia information, such as clinical notes and medical records, but also offer new possibilities for people to detect, monitor, and manage their health from anywhere at any time. Digital health technologies have the potential to improve patient care by making it more efficient, effective, and cost-effective. Utilizing digital devices and technologies can have a positive impact on many health conditions. This research focuses on dysphonia, a change in the sound of the voice that affects around one-third of individuals at some point in their lives. Voice disorders are becoming more common, despite being often overlooked. Mobile healthcare systems can provide quick and efficient assistance for detecting voice disorders. To make these systems reliable and accurate, it is important to develop an algorithm that can classify intelligently healthy and pathological voices. To achieve this task, we utilized a combination of several datasets such as Saarbruecken voice dataset (SVD), the Massachusetts Eye and Ear Infirmary database (MEEI), and a few private datasets of various voices (healthy and pathological) Additionally, we applied multiple machine learning algorithms, including decision tree, random forest, and support vector machine, to evaluate and determine the most effective algorithm among them for the detection of voice disorders. The experimental analyses are performed in terms of sensitivity, accuracy, receiver operating characteristic area, specificity, F-score and recall. The results demonstrated that the support vector machine algorithm, depending on the features selected by using appropriate feature selection methods, proved to be the most accurate in detecting voice diseases.

1. Introduction

The implementation of mobile devices for transmitting digital data or controlling and monitoring diseases has gained significant interest from both the research and business fields (Al-Dhief et al., 2020). These devices offer numerous benefits for developing efficient mobile health (mHealth) systems, enabling doctors and patients to access audio-visual notes, medical records and drug information (Sharma et al., 2022). mHealth technology can help identify and prevent diseases, facilitate decision-making, manage chronic emergencies, enhance the quality of patient care, and reduce healthcare costs.

Pathological conditions, including widespread cardiovascular diseases, can be detected and monitored using wearable sensors and wireless communication. The development of these monitoring systems has been made possible due to advancements in cloud technology and the Internet of Things (IoT) (Philip et al., 2021). These discreet, convenient, and user-friendly systems allow for the tracking and examination of health information, benefiting those with cardiovascular issues and supporting their physical therapy (Dias and Paulo Silva Cunha, 2018; Subramaniam et al., 2022). While health monitoring systems for cardiovascular diseases are widely recognized and appreciated, there are others less well-known and often undervalued. A voice disorder known as dysphonia affects the volume, sound, and pitch of the human voice. Approximately 10% of the population experiences it, mostly as a result of unhealthy social practices and overuse of the voice (Kirmayer

* Corresponding author.
E-mail address: mujeeb.rehman@dmu.ac.uk (M. Ur Rehman).

https://doi.org/10.1016/j.engappai.2024.108047
Received 7 August 2023; Received in revised form 2 December 2023; Accepted 4 February 2024
Available online 8 February 2024

et al., 2003). Despite their commonality, many individuals with voice disorders do not seek medical help. M-health solutions could offer an effective way to diagnose and screen for voice disorders (Al-Dhief et al., 2022).

The detection of voice pathology in a clinical setting is done through various procedures, one of which is acoustic analysis. This involves evaluating the voice signal to determine if there are any changes in the vocal tract, using specific parameters as outlined by the SIFEL (Società Italiana di Foniatria e Logopedia), protocol (Vernero and Schindler, 2012). The SIFEL protocol was created by the Italian Society of Logopedics and Phoniatrics based on guidelines from the European Society of Laryngology's Committee for Phoniatrics. Acoustic analysis is a non-invasive diagnostic tool used in conjunction with other medical exams such as the laryngoscopic examination, which involves visual inspection of the vocal folds (Angerstein et al., 2019).

Assessing voice health involves measuring various acoustic parameters (Paniagua et al., 2020). However, the accuracy of such parameters in identifying voice disorders often relies on the algorithms used for classification. Hence, researchers concentrate mainly on examining voice features and using classification methods for improved accuracy (Calvo and D'Mello, 2010; Mamyrbayev et al., 2019; Xu et al., 2020). Recently, speech pathology has shifted its focus to machine learning methods. AI-based extrapolation of voice disorders in speech and language pathology involves using artificial intelligence algorithms to analyze speech patterns and detect changes or abnormalities that may indicate a voice disorder. This can be done by analyzing features such as pitch, loudness, and voice quality and comparing them to a database of normal speech patterns. AI-based approaches have the potential to improve the accuracy and efficiency of voice disorder diagnosis in speech and language pathology (Idrisoglu et al., 2023). However, more research is needed to fully understand the potential of these methods and to ensure their accuracy and reliability.

Machine learning algorithms can objectively evaluate speech processing to identify pathological and healthy voices (AL-Dhief et al., 2020; Kim et al., 2020). These algorithms can greatly aid in the early detection of voice disorders or the assessment of voice quality before and after surgery. They offer techniques, methods, and tools that can assist in diagnosis across various medical fields. Speech analysis of voice pathology systems has employed various machine learning algorithms, including Decision Tree (DT) (Mohammed et al., 2021), Extreme Learning Machine (ELM) (Al-Dhief et al., 2021), Naïve Bayes (NB) and Support Vector Machine (SVM) (Selvakumari and Radha, 2017). Thus, these algorithms have been shown to be effective and efficient for the classification of pathological and healthy voices. However, some algorithms still experience challenges such as low accuracy in classification, prolonged processing, or heavy resource utilization in voice pathology identification and monitoring systems (Mittal and Sharma, 2021; Muhammad and Melhem, 2014; Geng et al., 2022; Al Mojaly et al., 2014). Moreover, several machine learning-based techniques require the entire dataset to be retrained when new data becomes available, resulting in longer processing times. This problem is a significant challenge as it leads to prolonged wait times for results. Another challenge is that a lot of research in voice disorder systems is limited to small datasets and only focuses on specific vowels such as /a/, ignoring others such as /i/ and /u/, as well as sentences. The shortcomings of voice disorder identification and detection systems can be summarized as follows:

- The previous studies primarily focused on binary classification, specifically distinguishing between normal voices and a single specific disorder, such as dysphonic voices or another isolated condition. Applying such existing schemes in the practical system would not be helpful for those patients with vocal laryngitis, puberphonia and cord paralysis. However, in the proposed study, a binary classification task is performed between normal and abnormal voices, whether it is cord paralysis, vocal fold granuloma, functional dysphonia, or any other voice disorder. Therefore, practically, the proposed method can be more suitable for classifying normal voices and other voices with any disorder.
- The majority of the voice pathology systems are limited to processing a single voice data, usually the vowel /a/, and do not take into account other vowels or sentences.
- Many studies employ small datasets of both healthy and pathological voice signals, resulting in low accuracy and lengthy processing times. Small datasets often exhibit deficiencies in diversity, increasing the risk of overfitting within machine learning models. Moreover, smaller datasets tend to be more susceptible to noise or outlier presence. Additionally, the complexity of algorithms might encounter challenges in extracting substantial information and patterns, thereby resulting in decreased accuracy and extended processing times (Ezugwu et al., 2022; Harar et al., 2017; Verde et al., 2018).
- Classifiers based on machine learning still struggle with achieving high accuracy.
- Voice pathology, identification and detection systems are only evaluated in terms of accuracy, sensitivity and area under the curve (AUC).

Thus, developing a trustworthy voice pathology identification and detection system using machine learning is essential to overcome such vulnerabilities. The contributions of this research are as follows:

1. A machine learning-based classifying technique is proposed to identify and detect specific types of voice pathology such as dysphonia. This is done based on the features extracted from the different voice signals.
2. The proposed system incorporates both healthy and pathological voice samples from Saarbruecken voice dataset (SVD), the Massachusetts Eye and Ear Infirmary database (MEEI), and a primary dataset focusing on sentences and vowels /a/, /i/, and /u/ spoken at three distinct pitch levels.
3. There is a significant amount of both pathological and healthy voice samples utilized to train and evaluate the proposed system.
4. There are a number of different assessment metrics used in order to evaluate and demonstrate how effective the proposed system is.

The goal is to be able to accurately distinguish between pathological and healthy voices. Moreover, we also aim to analyze the recognition of voice disorders using patient information such as gender and age, along with various features extracted from voice signals. The parameters used in the analysis such as jitter, shimmer, time period Fundamental Frequency (F0), and periodicity (Jothilakshmi, 2014; Mohammed et al., 2020).

The novelty of the proposed work lies in several major things: The proposed model demonstrates the ability to not only detect the presence of pathology in the voice but also identify the specific type of pathology. This added capability sets it apart and enhances its utility compared to the existing models that only focus on the presence or absence of pathology. Moreover, the proposed model is capable of detecting specific types of voice disorders such as dysphonia. The primary dataset used in the study is collected from various hospitals and medical centers. This unique dataset is then combined with publicly available datasets. The integration of these diverse data sources significantly contributes to the enhanced accuracy of the proposed model compared to relying only on online datasets.

The rest of the paper is structured as follows: Section 2 reviews previous research on using machine learning for voice disorder identification. Section 3 provides the material and the methodology proposed in this study. Section 4 provides the details of the dataset, extracted features, feature selection methods and classifiers used in the proposed study. Moreover, the experimental results and analysis are also given in Section 4. Section 5 concludes the proposed research and provides some future recommendations that can be useful to improve the proposed work.

2. Related work

Speech, or the voice signal, has various uses, from recognizing emotions (Kwon, 2019) to determining a patient's healthcare status (Hossain, 2016). m-health solutions presented in Cesari et al. (2018), Seedat et al. (2020) and Spadaro et al. (2022) use voice signals to assess voice health, and systems use them to gauge emotional condition (García et al., 2019; Alaiad et al., 2019). Voice pathology detection often employs machine learning and various approaches have been developed in the past few years to increase accuracy in identifying and detecting pathological and healthy voices.

Recent research has aimed to improve accuracy and results in voice pathology detection by exploring various voice quality parameters, such as jitter, shimmer and fundamental frequency of a speech and a voice signal (Vizza et al., 2019; Brockmann et al., 2008). Additionally, several acoustic features have been studied in the past few decades, including the Mel-Frequency Cepstral Coefficient (MFCC) (Kelly and Gobl, 2011) and Multidimensional Voice Program (MDVP) (Nicastri et al., 2004), which are extracted from speech signals for analysis. Common classifiers used in these studies include Artificial Neural Network (ANN) (Souissi and Cherif, 2016), Hidden Markov Models (HMM) (Srivastava et al., 2022) and SVM (Reid et al., 2022). Researchers have utilized datasets such as the Saarbrücken Voice Database (SVD) (Reid et al., 2022), Arabic Voice Pathology Database (AVPD) (Gidaye et al., 2022) and Massachusetts Eye and Ear Infirmary Database (MEEI) (Zhou et al., 2022) for the detection and identification of voice disorders. Despite the fact that there has been progress made in this field, the research is still in the preliminary phases and needs more investigation, employing a larger variety of speech samples and datasets as well as other machine learning algorithms.

The existing studies aim to identify parameters for measuring voice quality and develop new classification systems for detecting voice disorders. In the field of voice signal processing, a machine learning algorithm known as SVM has been commonly used in recent years.

In Godino-Llorente et al. (2005), Godino et al. used SVM to classify the pathological and healthy voices with 95% accuracy. Moreover, the dataset used in their study was limited. It contained only 173 pathological and 53 healthy voice samples. The specific pathologies of these voices are also not provided in their study.

In Yang et al. (2014), SVM was also utilized to determine the presence of dysphonia, examining four types of pathology. The study used MFCC and LDA for dimensionality reduction, resulting in an accuracy of 86% in identifying a pathology. However, the dataset was limited, consisting of only 70 pathological and 40 healthy voices from the Saarbruechen Voice Database.

In Al-Dhief et al. (2021), Dhief et al. developed a voice pathology detection system using voice signals from the Saarbrücken voice database (/a/vowel). Mel-Frequency Cepstral Coefficients (MFCC) extract signal features, and Support Vector Machines (SVM) classify healthy and pathological voices. The evaluation shows SVM achieves 84.37% accuracy, 90.90% specificity, and 80.95% sensitivity, indicating strong potential for identifying voice disorders.

MFCC parameters were evaluated in several other studies, including Cordeiro et al. (2015) and Amara et al. (2016). In Cordeiro et al. (2015), Cordeiro et al. analyzed subjects with nodules, edema, and unilateral vocal fold paralysis, with a less-than-optimal accuracy of 77.90%. In Amara et al. (2016), Amara, et al. studied spasmodic dysphonia and used to diagnose patients using machine learning algorithms, but algorithms such as SVM and Gaussian Mixture Models (GMM) were evaluated using limited voice samples from the MEEI database, including healthy and pathological ones.

By evaluating MFCC, shimmer and jitter, El Emary et al. (2014) classified speech and voice signals. They used the GMM algorithm to detect neurological voice disorders on a small dataset of 38 pathological and 63 healthy voices from the SVD database. In Fonseca et al. (2007), Fonseca et al. employed an algorithm based on LS-SVM and linear prediction coefficients to detect laryngeal voice disorder and tested it on a private dataset.

In Barreira and Ling (2020), Barreira et al. presented a novel speech pathology diagnosis system. Using Kullback–Leibler Divergence (KLD) on histograms of voice and speech signals, the system generates two modified spectra known as Higher Amplitude Suppression Spectrum (HASS-KLD) and H-KLD (Barreira and Ling, 2020). These spectra quantify the distribution of voice and speech frames and offer high accuracy with limited parameters. H-KLD assesses the difference between the probability distributions of the voice and speech frames, while HASS-KLD evaluates the dynamic features of the voice and speech signals. The extracted features are then input into two classifiers. The system was tested using the MEEI dataset which includes 53 normal and 173 pathological voice samples achieving an accuracy of 99.55%. However, the process is time-intensive, utilizing multiple classifiers and feature extraction techniques.

In Wang and Jo (2007), Wang et al. used a private dataset collected at the Busan National University Hospital. The authors used HMM, GMM, and SVM to classify pathological voices, including vocal cord palsy, vocal polyps, edema, glottic cancer, and laryngitis. Many studies in the literature use private datasets, such as Ritchings et al. (2002) and Boyanov and Hadjitodorov (1997). Without defining the specific pathologies, Ritchings et al. (2002) used 77 pathological voices from a private dataset collected from the Christie Hospitals in Manchester to train and test their proposed system. While in Boyanov and Hadjitodorov (1997), Boyanov et al. gathered data from the University Hospital of Sofia and utilized it in combination with the K-nearest neighbors algorithm and linear discriminant analysis to diagnose laryngeal pathology.

In Fonseca et al. (2020), Fonseca et al. proposed a system for voice pathology identification and detection that uses three different extractions, such as zero-crossing rate, signal entropy and energy. The Saarbrücken Voice Database (SVD) was used, and the maximum accuracy was 95%. Another work extracted glottal signal parameters for voice disorder detection and applied k-NN and SVM to classify the voice signal using SVD. SVM had an accuracy of 98.5% while k-NN had 88.2%, but with a limited set of voice samples.

Alhussein and Muhammad (2019) proposed a system for voice pathology detection using a mobile platform and smart healthcare framework that is based on a deep learning model known as Convolutional Neural Network (CNN). Voice signals are recorded on smartphones, processed and analyzed in the cloud, and classified into three different parallel models. Parallel CNNs achieved 95.5% accuracy on the SVD dataset with 686 healthy voice samples and 1342 pathological voice samples, but only for sustained vowel /a/ spoken at normal pitch.

In Upadhya and Cheeran (2018), Upadhya et al. proposed a system to classify healthy and pathological voices that incorporates phonation features such as fundamental frequency of pitch, energy and stability, as well as cepstral features, including MFCC. A neural network having three different layers is used as the classifier that achieved an accuracy of 95.6% for phonation features and 81.1% accuracy for cepstral features. However, due to the limited number of samples, the performance may decrease if tested on a larger database.

In Chen et al. (2023), Chen et al. conducted a research where they obtained recordings of sustained phonations of the vowels /a/ and /i/ from a clinical database. The dataset consisted of 238 individuals with dysphonia and 223 healthy voices. To ensure consistency, all audio clips were divided into multiple 1.5-second segments and normalized for loudness. These segments were then utilized as input features for a convolutional neural network (CNN) to perform a binary classification task. The most favorable results were obtained when classifications were based on all segments of both vowels, achieving an impressive accuracy of 95%.

In Fang et al. (2019), Fang et al. collected voice samples, comprising 60 normal voice samples and 402 pathological voice samples from individuals diagnosed with eight common clinical voice disorders.

To assess the performance of different machine learning algorithms, namely the deep neural network (DNN) and support vector machine, a fivefold cross-validation approach was employed. The accuracy of the DNN algorithm in detecting voice disorders was found to be 94.26% for male subjects and 90.52% for female subjects.

In Powell et al. (2019), Powell et al. proposed a study in which they obtained acoustic recordings from a clinical database. The dataset comprised recordings from 10 vocally healthy speakers and 70 patients diagnosed with one of seven different voice disorders (with 10 patients per diagnosis). To process the acoustic signals, spectrograms were generated and used as input for a convolutional neural network (CNN) developed with the Keras library. To assess the performance of the models, a 10-fold cross-validation technique was employed for validation. The binary classification accuracies varied across the different diagnostic categories, ranging from 58% to 90%.

In Al-Dhief et al. (2021), Laverde et al. proposed a voice pathology detection and identification system using ANN and SVM classifiers and Particle Swarm Optimization for optimal parameters. Three types of features such as noise features, common voice features and acoustic features are extracted from each voice sample including healthy and pathological. The voice dataset (SVD) is divided into three groups (D1, D2, and D3), with each group containing the same number of voice samples. D1 consists of normal-pitched vowel /a/ sounds; D2 contains sentences; and D3 holds recorded sentences. The SVM provides an accuracy of 92.77%, while the ANN has a 93.27% accuracy, both based on group D3 of recorded sentences. However, the system's performance for other vowels such as /u/ and /i/ pronounced with varied intonations was not assessed using the speech database (SVD).

A system based on LP analysis for classifying healthy and disordered voice samples is proposed in Ali et al. (2017). The voice tract is divided into multiple tubes using LP analysis. The MEEI dataset was used, which contains 173 pathological and 53 healthy samples. The GMM algorithm is used with an increasing number of Gaussian mixtures (4 to 50) to classify the voice signals with an accuracy of 99.94% for the vowel /a/ and 99.75% for recorded sentences. However, similar to the work proposed in Al-Dhief et al. (2021), the system is only evaluated on vowel /a/ and has limited normal sample data.

The summary of existing studies on voice disorder detection is outlined in Table 1, encompassing strengths, vulnerabilities, and potential solutions.

3. Material and methods

The proposed research evaluated the accuracy of various machine learning algorithms to classify healthy and pathological voices, with the goal of finding the most reliable algorithm. The selected algorithm will be used in the proposed m-healthcare system, which will allow speech signals to be recorded using a smartphone, tablet, or other smart device equipped with a voice recorder. The features such as jitter, shimmer, and fundamental frequency will be extracted and analyzed by the classifier to determine if the patient has a voice disorder, as displayed in Fig. 1.

Specifically, the SVM which is the most commonly used algorithm in the literature due to the kernel function is used, as well as a few other classifiers for detecting voice disorders are also incorporated to gauge the performance of each classifier. The performance and experimental analysis are carried out using WEKA (Arora, 2012), a widely used tool for data mining that was chosen for its affordability, versatility, and efficiency in data analysis.

The following sections introduce the dataset utilized in the proposed research, the features that are extracted from the speech and voice signal for classification purposes, and the machine learning algorithm used for comparison purposes.

3.1. Dataset used in the proposed research

The proposed research involves a dataset that is made up of a combination of various existing datasets such as MEEI (El Emary et al., 2014), SVD (Souissi and Cherif, 2015), and the private dataset. The aim of combining multiple datasets is to increase the overall size of the dataset for improved training. In addition to the existing datasets, we also gathered a dataset that consists of 2015 healthy voice samples and 3678 pathological voice samples. This combined dataset, referred to as the "Collected and Multiple Existing Dataset (CMED)", includes 2857 healthy voice and speech samples and 5301 pathological voice samples. Each sample is taken from a different speaker, ensuring that no speaker contributes more than one sample to the dataset. The voice samples are stored in the ".wav" format, representing the format used for each voice signal recording.

To avoid biases and maintain consistent class distribution across various data subsets, especially when dividing data into training and testing sets in machine learning, a commonly employed method called stratification is utilized. The implementation of this technique follows Algorithm 1.

Algorithm 1 Process for the implementation of stratification

```
Start
Input Different datasets: (SVD, MEEI, and private dataset)
Define Features and labels: Features: X, Labels: Y
classDistribution = tabulate(Y)        ▷ Identify distribution of classes
Split dataset into test_dataset and train_dataset
testdataset = 0.2; traindataset = 0.8
[trainInd, testInd] = splitDataWithStratification(X, Y, testdataset)
function [trainInd, testInd] = splitDataWithStratification(X, Y, testRatio)
                                       ▷ Spliting data while maintaining class proportions
    classes = unique(Y);
    trainInd = [ ];
    testInd = [ ];
    for i = 1:numel(classes)
        classIdx = find(Y == classes(i));
        n = numel(classIdx);
        nTest = round(n * testRatio);
        shuffledIdx = classIdx(randperm(n));        ▷ Randomly shuffle indices for each class
        trainInd = [trainInd; shuffledIdx(nTest + 1:end)];
        testInd = [testInd; shuffledIdx(1:nTest)];
    end
end
End
```

This dataset includes voice samples of the vowels /a/, /i/, and /u/. It is preferable to use vowels for evaluating the patient's voice quality because it eliminates language-related issues and is common in voice disorder detection. In the clinical context, the vowel /a/ is often used for detecting and identifying voice disorders.

In the case of the SVD dataset, the SVD is primarily used for speech and voice analysis. It contains recordings of various speech and voice samples rather than reports of diagnosed voice disorders. Therefore, specific voice disorders reported within the SVD may not be explicitly documented. However, in the proposed research this database is used to study aspects of voice quality, and speech pathology, which may indirectly contribute to understanding voice disorders like dysphonia, vocal nodules, and other speech impairments.

Fig. 2 shows the healthy and pathological voice and speech signals. The healthy signals in Fig. 2(a, c, and e) have a repetitive pattern or periodicity, indicating a healthy voice. Whereas, the pathological signals in Fig. 2(b, d, and f) have a high degree of unpredictability or randomness, indicating they the patient is suffering from any voice disorder. The proposed model identifies pathological voices with a

Table 1
Summary of existing studies on voice disorder detection.

Methodology name	Application domain	Advantages	Vulnerabilities	Potential solutions
SVM (Godino-Llorente et al., 2005)	Pathological and healthy voice classification	High accuracy (95%)	Limited dataset, lack of pathology specifies	Utilize larger and diverse datasets
KLD-based system (Barreira and Ling, 2020)	Speech Pathology diagnosis	High accuracy	Time-intensive process	Optimize process, streamline technique
SVM, MFCC, /a/ vowel (Al-Dhief et al., 2021)	Voice pathology detection	Strong accuracy, specificity, and sensitivity	Dataset limited to specific vowel	Explore diverse vowels, augment dataset
ANN, SVM PSO (Al-Dhief et al., 2021)	Pathology detection using ANN and SVM	Voice High accuracy	Limited vowels tested	Explore diverse vowels, augment dataset
CNN-based study (Chen et al., 2023)	Dysphonia detection using CNN	Good accuracy	Limited dataset	Augment dataset, explore more diverse data
Acoustic recordings, CNN (Powell et al., 2019)	Voice disorder classification using CNN	Algorithm based classification	Varied accuracies	Optimize model, explore diverse datasets
ML algorithms (Fang et al., 2019)	Voice disorder detection using ML	High accuracy	Limited dataset usage	Expand dataset diversity
Voice analysis (Cordeiro et al., 2015)	Analysis of nodules, edema, vocal fold paralysis	Study domain-specific voice disorder	Lower accuracy	Investigate new features, expand dataset
GMM, SVD database (El Emary et al., 2014)	Neurological disorder detection	ML-based neurological voice disorder detection	Small dataset	Augment dataset, explore new algorithms
SVM, MFCC, LDA (Yang et al., 2014)	Dysphonia identification	Good accuracy (86%)	Small dataset	Expand dataset, explore more features

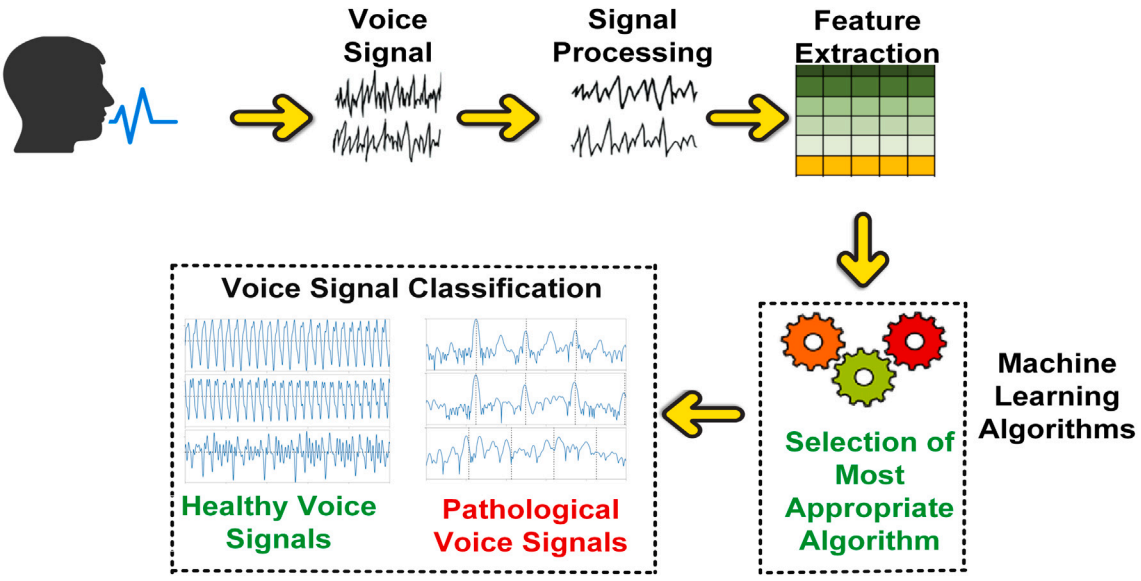


Fig. 1. Classification of normal and pathological voices.

severity level categorized as moderate or severe, based on the statistical values of individual features.

For the experiments, we selected a total of 5693 samples, including both healthy and pathological voices. The distribution of the data samples in each dataset is shown below.

For MEEI dataset:

- Pathological voice samples = 173; Male = 51; Female = 122
- Healthy voice samples = 53; Male = 38; Female = 15

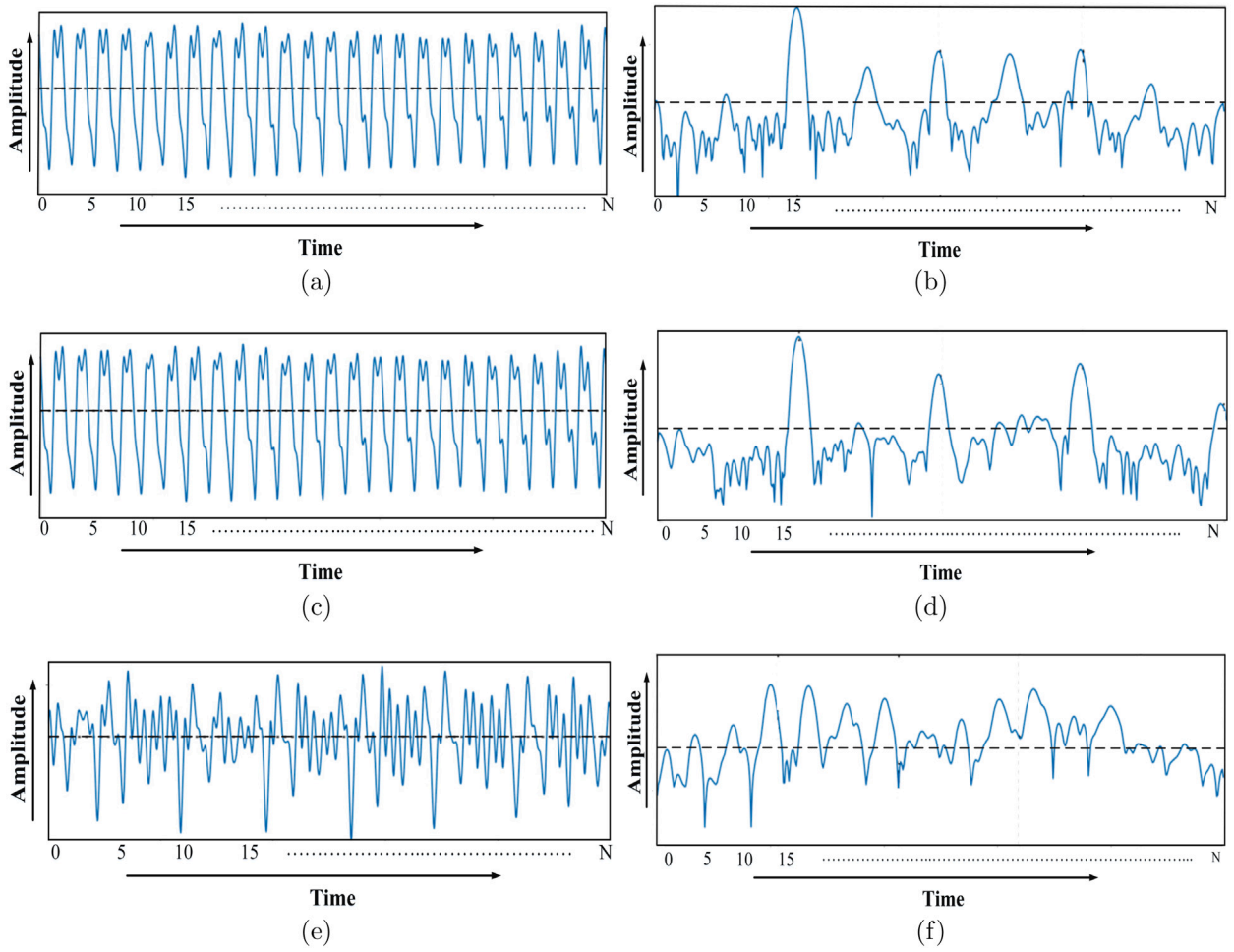


Fig. 2. Healthy and pathological voice and speech signals.

For SVD dataset:

- Pathological voice samples = 1342; Male = 563; Female = 1342
- Healthy voice samples = 686; Male = 518; Female = 168

Primary dataset:

- Pathological voice samples = 70; Male = 59; Female = 31
- Healthy voice samples = 40; Male = 16; Female = 24

Private dataset used in El Emary et al. (2014):

- Pathological voice samples = 38; Male = 12; Female = 26
- Healthy voice samples = 63; Male = 34; Female = 29

For CMED dataset:

- Pathological voice samples = 5301; Male =2788; Female = 2513
- Healthy voice samples = 2857; Male =1990; Female = 867

Table 2 shows details about the chosen samples, including the number and percentage (up to two digits after decimal) of voices for each age and gender. There are fewer female samples due to the higher occurrence of voice disorders in male subjects. Different researchers have mentioned the labels of the samples in the dataset to ensure they are all named or labeled the same way (Chaiani et al., 2022; Fang et al., 2019). Additionally, a portion of the dataset (primary dataset) is also collected from various hospitals and medical centers. The collected dataset is then merged with existing datasets such as SVD, MEEI and dataset used in El Emary et al. (2014) to create a new dataset for voice pathology detection. All datasets used in the proposed research utilized voice recordings across various pitch levels—normal, high, or low. Furthermore, the features extracted from each voice concern sustained vowels, aiding in the detection of normal, high, and low-pitch voices. The differentiation depends on the statistical values of the features that are extracted from these different recorded voices.

The samples within the collected dataset are properly labeled by doctors and medical practitioners, ensuring accurate and reliable annotations. All available pathological and healthy voices from the CMED (SVD + MEEI + dataset used in El Emary et al. (2014) + primary dataset) are used. Moreover, the flow diagram of the proposed work for the classification of healthy and pathological voices is shown in Fig. 3.

3.2. Data pre-processing

To enhance the quality of voice recording, a few processing techniques are applied to each recording. The details of the processing are given in the following steps:

- **Short-Time Fourier Transform (STFT):** The voice recordings are divided into small frames of length N samples. After the division into short frames, the short-time Fourier transform (STFT) is applied to each short frame to convert the signal into the frequency domain from the time domain. For a given frame $x[n]$ of length N, the STFT is calculated using Eq. (1):

$$X(k,m) = STFT\{x[n]\} = \sum_{n=0}^{N-1} x[n] \cdot w[n-mH] \cdot e^{-j2/N} \quad (1)$$

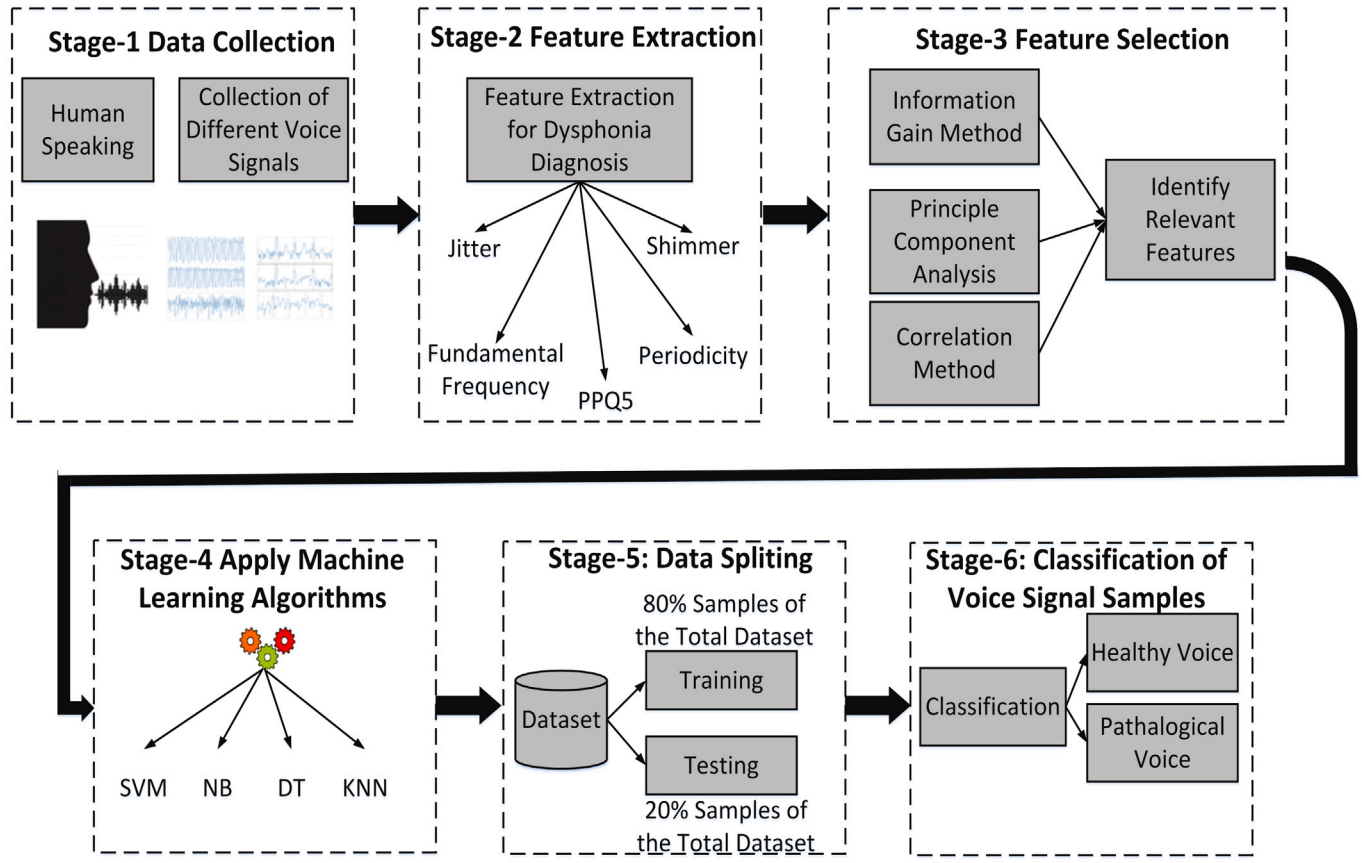


Fig. 3. Flow diagram of the proposed work.

Table 2
Healthy and pathological voice samples in the CMED dataset.

Voice samples	Age bracket	Gender	No. of samples (NoS)	%age of NoS
Pathological	15–30	Female	267	3.28%
	31–40	Female	286	3.50%
	41–50	Female	469	5.74%
	51–60	Female	897	10.99%
	61–70	Female	594	7.28%
Pathological	15–30	Male	82	1.00%
	31–40	Male	364	4.46%
	41–50	Male	425	5.20%
	51–60	Male	794	9.73%
	61–70	Male	1123	13.76%
Healthy	15–30	Female	631	7.73%
	31–40	Female	539	6.60%
	41–50	Female	330	4.04%
	51–60	Female	214	2.62%
	61–70	Female	276	3.38%
Healthy	15–30	Male	280	3.43%
	31–40	Male	251	3.07%
	41–50	Male	149	1.82%
	51–60	Male	115	1.40%
	61–70+	Male	72	0.88%
Total (T_1)	15–70+	Female	4503	55.19%
Total (T_2)	15–70+	Male	3655	44.80%
Total (T_t)	15–70+	T_1 + T_2	8158	100%
Voice samples (a)	–	Male+Female	4610	56.50%
Voice samples (i)	–	Male+Female	2561	31.39%
Voice samples (u)	–	Male+Female	987	12.09%

where, $X(k,m)$ is the complex value of the kth frequency bin at frame m. $w[n-mH]$ is the window function applied to the frame, H is the hop size, k ranges from 0 to N–1 and m represented the frame index.

- **Noise Estimation:** Now estimate the level of noise by computing the average for the magnitude spectrum of the noisy frames as $|N(k)|$. Then calculate the magnitude of the noisy version of the signal's spectrum $|Y(k,m)|$ which is obtained from STFT.
- **Spectral Subtraction:** Subtract the estimated noise spectrum from the noisy signal spectrum using Eq. (2).

$$|X_{clean}(k,m)| = max(|Y(k,m)| - \alpha|N(k)|, 0) \tag{2}$$

where, $|X_{clean}(k,m)|$ denotes the magnitude of the spectrum of the clean signal, and α is the scaling factor which is set to 1.9 for controlling the amount of noise reduction.

- **Inverse STFT:** Apply inverse STFT (ISTFT) using Eq. (3) to obtain the clean time-domain signal.

$$x_{clean}[n] = ISFIT\{X_{clean}(k,m)\} = \sum_{m=0}^{M-1}\left(X_{clean}(k,m)\cdot e^{j2\pi kn/N}\right)\cdot w'[n-mH] \tag{3}$$

where, $x_{clean}[n]$ is the cleaned signal and $w'[n-mH]$ is the inverse window function.

Repeat each step for every voice signal to acquire a clean and enhanced version of the voice recordings.

3.3. Features used in the proposed research

Feature extraction is crucial for enhancing the experimental analysis and classification. The selection of speech signal features for the proposed study is based on two categories: (A) the key parameters utilized by experts in clinical evaluations and (B) the main features

Table 3
Summary of the features used in the proposed work.

Features	Mathematical equations	Explanation
Jitter (absolute)	$jitter(absolute) = \frac{1}{P-1}\sum_{i=1}^{P-1}(\lvert T_i - T_{i-1}\rvert)$	T_i: length of each period in seconds, P: total number of periods
Local jitter	$jitter(local) = \frac{\frac{1}{P-1}\sum_{i=1}^{P-1}(\lvert T_i - T_{i-1}\rvert)}{\frac{1}{P}\sum_{i=1}^{P}(T_i)}$	//
Relative Average Perturbation (RAP) Jitter	$RAP(jitter) = \frac{\frac{1}{P-1}\sum_{i=1}^{P-1}(\lvert T_i - (\frac{1}{3}\sum_{p=i-1}^{i+1}(T_p))\rvert)}{\frac{1}{P}\sum_{i=1}^{P}(T_i)}$	//
Five-Point Period Perturbation Quotient (PPQ5)	$PPQ5 = \frac{\frac{1}{P-1}\sum_{i=1}^{P-1}(\lvert T_i - (\frac{1}{5}\sum_{p=i-1}^{i+1}(T_p))\rvert)}{\frac{1}{P}\sum_{i=1}^{P}(T_i)}$	//
Shimmer (absolute)	Shimmer (absolute) $= \frac{1}{P-1}\sum_{i=1}^{P-1}(\lvert 20 * log(\frac{A_{i+1}}{A_i})\rvert)$	//
Shimmer (local)	Shimmer (local) $= \frac{\frac{1}{P-1}\sum_{i=1}^{P-1}(\lvert A_i - A_{i+1}\rvert)}{\frac{1}{P}\sum_{i=1}^{P}(A_i)} \times 100$	A_i: Amplitude of the recorded voice signal
Shimmer (APQ3)	Shimmer (APQ3) $= \frac{\frac{1}{P-1}\sum_{i=1}^{P-1}(\lvert A_i - (\frac{1}{3}\sum_{p=i-1}^{i+1}(A_p))\rvert)}{\frac{1}{P}\sum_{i=1}^{P}(A_i)}$	//
Shimmer (APQ5)	Shimmer (APQ5) $= \frac{\frac{1}{P-1}\sum_{i=2}^{P-2}(\lvert A_i - (\frac{1}{5}\sum_{p=i-2}^{i+2}(A_p))\rvert)}{\frac{1}{P}\sum_{i=1}^{P}(A_i)}$	//
Other features →	Fundamental frequency, periodicity	–

commonly employed in the existing studies that apply machine learning algorithms for the classification of healthy and pathological voices.

Instead of using dynamic range as input to the AI/ML model, various technical aspects such as Frequency, Periodicity, Jitter (absolute), Local jitter, Relative Average Perturbation (RAP) Jitter, Five-Point Period Perturbation Quotient (PPQ5), Shimmer (absolute), Shimmer (local), Shimmer (APQ3), and Shimmer (APQ5) are employed to determine and compare the accuracy of the proposed model.

A detailed explanation of such features can be found in Shafique et al. (2021) and Li et al. (2007). The summary of the features used in the proposed work is provided in Table 3.

Each item (i), in the dataset used for the proposed research consists of the following data:

- **Patient ID:** An alphanumeric value serves as the identifier for the patient.
- **Age:** Measured in years
- **Features:** F0, periodicity, Jitter (absolute), Local jitter, Relative Average Perturbation (RAP) Jitter, Five-Point Period Perturbation Quotient (PPQ5), Shimmer (absolute), Shimmer (local), Shimmer (APQ3) and Shimmer (APQ5).
- **Class/Target value:** Healthy voice and Pathological voice

The choice of feature extraction techniques, namely fundamental frequency, periodicity, various jitter measurements (absolute, local, RAP, PPQ5), as well as shimmer measurements (absolute, local, APQ3, APQ5), is founded on extensive research in the field of voice disorder detection and voice pathology analysis. These features have been widely utilized in existing literature (Mesallam et al., 2017; Mamun et al., 2022; Chaiani et al., 2022) for their effectiveness in capturing important aspects of voice quality and characteristics associated with pathological conditions. Moreover, these features ensure the comprehensive coverage of important parameters for accurate detection and classification of voice pathology.

Apart from the features used in the proposed research, the determination of the scale of vocal variations relies on factors such as the qualities of acoustic transducers and voice recording devices. It necessitates a comprehensive analysis to accurately assess vocal variations. This analysis involves extracting essential features like frequency and shimmer from each speech signal within the dataset. The speech samples in the dataset are not recorded using any particular microphone or acoustic transducer. However, the transducers can have a significant impact on voice and speech. Therefore, determining the precise scale or level of such vocal variations that are only based on acoustic transducers would require further analysis and context-specific information. The proposed model in this research only works on speech samples not recorded by any acoustic transducer. An acoustic transducer refers to any device like a microphone. However, the proposed model specifically works with human voice recordings that are not captured using a microphone. The aim is to exclude any potential inaccuracies or disturbances caused by microphone-related noise or faults. Therefore, the model only utilizes voices directly sourced from humans to ensure accuracy and reliability in its construction.

3.4. Classifiers used in the proposed study

To conduct a comprehensive comparison, we chose various machine learning algorithms to serve as representatives of the target class with similar features. These algorithms are:

3.4.1. Support vector machine (SVM)

SVM is a supervised machine learning algorithm used for classification and regression analysis. It is based on the concept of finding the hyperplane (a line or a higher-dimensional plane) that maximally separates the classes in the training data.

In a two-class problem, the SVM algorithm finds the hyperplane that separates the two classes with the largest margin, known as the maximum margin classifier. This hyperplane is called the maximum margin hyperplane, and the points closest to the hyperplane on either side are called support vectors. The margin is the space between the hyperplane and the data points that are closest to it.

To classify data, multiple inputs or features must be present. The dimensions of the dataset are determined by the number of features used. If a dataset has 10 features, it is considered 10-dimensional. it can be expressed as :

For Multiple (N) dimensional dataset: $Z = D_1, D_2, D_3, \ldots, D_N$

where $D_1, D_2, D_3, \ldots, D_N$ are the independent variables of features and Z is the output variable or target class.

When creating a dataset, the number of features and the number of output categories do not have to match. It depends on the desired number of categories the data points should be classified into. SVM utilizes a

line or hyperplane to classify the data. In a 2D dataset, a line (called the support vector) is used to categorize the data with the largest margins. However, for higher-dimensional datasets, a hyperplane is used instead which can be expressed as:

$$AX + Y = 0$$

Whereas the bias Y and the input feature vector X are related by the vector A, which has the same number of dimensions as X. In the proposed work, an 11-dimensional dataset is used, so the product of A and X can be represented as AX.

$$A^{1*}x^1 + A^{2*}x^2 + A^{3*}x^3 + A^{4*}x^4 + A^{5*}x^5 + A^{6*}x^6 + A^{7*}x^7$$

While predicting the target class, the following equation can be used.

$$Z = sign(AX + Y) \tag{4}$$

The sign of input depends on whether it is positive or negative. If it is positive, the sign returns + 1, and if it is negative, the sign returns − 1. The input consists of a feature vector X_i and a label Z_i which can either be +1 or −1. This can be expressed as:

$$\begin{cases} AX - Y \geq +1 & Z_i = +1 \\ AX - Y \leq -1 & Z_i = -1 \end{cases} \tag{5}$$

SVM also utilizes various types of kernels, such as radial basis function (RBF) or polynomials. In the proposed work, such kernels are used in the implementation of SVM and the results are reported in Section 4.1. The other classifiers used in the proposed research are decision tree, Naive byes, and K-nearest neighbor. The details of such classifiers can be found in Myles et al. (2004), Reddy et al. (2022) and Bhowmik et al. (2022).

3.4.2. Selection of hyperparameters

The details of hyperparameters for DT, NB, RF and KNN are given below:

1. Decision Trees (DT):
   - max_depth: A positive integer, i.e., 5.
   - min_samples_split: a small integer i.e., 2.
   - min_samples_leaf: a small integer, i.e., 5.
   - criterion: Gini
2. Naive Bayes (NB):
   - Smoothing parameter: A small positive value, i.e., 0.1.
3. Random Forest (RF):
   - n_estimators: A positive integer i.e., 50.
   - max_depth: A positive integer i.e., 76.
   - min_samples_split: A small integer., i.e., 5.
   - min_samples_leaf: A small integer, i.e., 2.
   - criterion: Gini
4. k-Nearest Neighbors (KNN):
   - n_neighbors: An odd positive integer, i.e., 5
   - weights: distance rather uniform
   - p: Set to 2 for Euclidean distance.

3.5. Evaluation metrics

The effectiveness of the proposed machine learning-based m-health classification system is assessed using performance metrics such as accuracy, sensitivity, specificity, and ROC area. Such parameters depend on the following measures.

- **True Positives** (τP): When the voice sample falls into the category of "Pathological voice" and the algorithm also recognizes that it is a Pathological voice.
- **True Negatives** (τN): When the voice sample falls into the category of "healthy voice" and the algorithm also recognizes that it is a healthy voice sample.
- **False Positives** ($_FP$): When the voice sample falls into the category of "pathological voice" and the algorithm recognizes that it is a healthy voice sample.
- **False Negatives** ($_FN$): When the voice sample falls into the category of "healthy voice" and the algorithm recognizes that it is a pathological voice sample.

Accuracy: Accuracy refers to the fraction of predictions made by a model that is correct. The mathematical formulation to find the accuracy of the model is given in Eq. (6) (Albadr et al., 2021).

$$Accuracy = \frac{\tau P + \tau N}{\tau P + \tau N + {}_FP + {}_FN} \tag{6}$$

Sensitivity: Sensitivity refers to the ability of a model to correctly classify positive examples (i.e., instances where the target variable is positive). Mathematically, it can be express using Eq. (7) (Albadr et al., 2022c):

$$Sensitivity = \frac{\tau P}{\tau P + {}_FN} \tag{7}$$

Specificity: Specificity in machine learning refers to the ability of a model to correctly identify negative cases, i.e. samples that belong to the non-target class. Mathematically, it can be given represented using Eq. (8): (Albadr et al., 2022b)

$$Specificity = \frac{\tau N}{\tau N + {}_FP} \tag{8}$$

Recall: Recall measures the ability of a classifier to correctly identify positive instances among all actual positive instances in a dataset. Mathematically it can be expressed using Eq. (9) (Albadr et al., 2022a).

$$Recall = \frac{\tau P}{\tau P + {}_FN} \tag{9}$$

F-score: F1-Score is a measure of a model's accuracy that balances precision and recall. It can be calculated using Eq. (10) (Albadr et al., 2023):

$$F - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

Receiver Operating Characteristic (ROC): The ROC area measures the accuracy of a classification algorithm. This metric plots the sensitivity of the algorithm against the inverse of specificity (1-specificity) and calculates the area under the curve (AUC). The AUC value can be thought of as the average sensitivity score across all specificity levels. A perfect classification will result in an AUC of 1, while an AUC of 0 means that the algorithm is completely inaccurate and misclassifies all positive cases as negative and vice versa.

4. Experimental results and discussion

All the steps performed in the development of the proposed model follow a Python 3.7 based implementation, employing pandas for data pre-processing and loading, while scikit-learn facilitates classifier implementation. Experimentation is conducted on a system equipped with an Intel i7 processor 12 GB RAM and Windows 11.

In our experiments, we employed cross-validation to avoid overfitting and enhance the unreliability of our predictions. Specifically, we utilized 10-fold cross-validation, which involved dividing the training dataset into 10 smaller subsets. For each of these 10 folds, a model is trained using k-1 folds as the training data and then evaluated on the remaining part of the dataset. During the development of the proposed model, 20% percent of the dataset comprises an equal distribution: half of it consists of samples from normal patients, while the other half contains samples of pathological voices. Each sample corresponds to a different speaker.

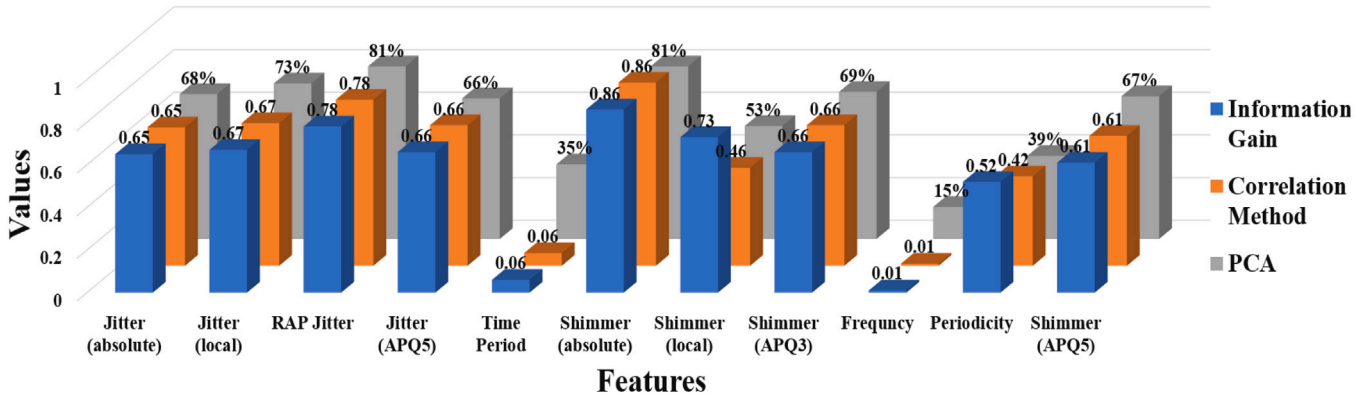


Fig. 4. Feature selection using information gain, correlation method and PCA.

Table 4
Statistical results obtained when incorporating SVM with RBF and polynomial kernel for different values of q and γ (Considering all features).

q (RBF)	q = 0.4	q = 0.8	q = 1.2	q = 1.6	q = 2.0	q = 2.4	q = 2.8	q = 3.2	q = 3.6	Mean	S.D
Accuracy (%)	90.32	89.9	88.76	90.25	90.65	91.67	88.61	87.46	85.66	89.25	1.73
Sensitivity (%)	89.66	88.45	90.13	90.66	90.45	91.73	90.56	90.36	86.99	89.88	1.31
Specificity (%)	86.92	87.65	89.99	90.12	91.56	90.24	90.46	91.56	88.99	89.72	1.50
F score	0.88	0.95	0.903	0.86	0.88	0.91	0.90	0.89	0.90	0.89	0.02
Recall	0.89	0.84	0.83	0.88	0.89	0.80	0.90	0.91	0.90	0.87	0.03
ROC area	0.86	0.85	0.88	0.89	0.90	0.88	0.89	0.87	0.90	0.88	0.01
q (polynomial)	q = 0.4	q = 0.8	q = 1.2	q = 1.6	q = 2.0	q = 2.4	q = 2.8	q = 3.2	q = 3.6	Mean	SD
Accuracy (%)	90.36	91.23	90.65	90.23	91.03	90.66	90.33	93.64	90.56	90.96	0.99
Sensitivity (%)	89.69	90.23	90.25	91.035	91.25	90.44	89.65	89.99	90.55	90.34	0.51
Specificity (%)	90.55	91.26	92.54	90.66	89.56	90.55	90.78	90.66	90.65	90.80	0.74
F score	0.89	0.91	0.92	0.90	0.92	0.93	0.90	0.89	0.91	0.90	0.01
Recall	0.90	0.89	0.89	0.91	0.90	0.89	0.88	0.91	0.90	0.89	0.01
ROC area	0.91	0.89	0.88	0.87	0.89	0.90	0.91	0.90	0.92	0.89	0.01
γ (RBF)	γ = 0.4	γ = 0.8	γ = 1.2	γ = 1.6	γ = 2.0	γ = 2.4	γ = 2.8	γ = 3.2	γ = 3.6	Mean	SD
Accuracy	90.66	90.12	90.21	90.65	91.06	90.56	92.56	91.46	90.57	90.87	0.70
Sensitivity	90.23	89.89	89.99	90.56	91.23	91.56	91.65	92.03	91.05	90.91	0.73
Specificity	88.98	89.68	90.89	89.99	90.56	90.36	91.56	90.32	92.55	90.54	0.98
F score	0.88	0.86	0.87	0.89	0.90	0.89	0.90	0.91	0.90	0.88	0.01
Recall	0.90	0.91	0.89	0.89	0.90	0.91	0.92	0.93	0.93	0.90	0.01
ROC area	0.88	0.87	0.90	0.91	0.90	0.91	0.90	0.89	0.88	0.89	0.01
γ (polynomial)	γ = 0.4	γ = 0.8	γ = 1.2	γ = 1.6	γ = 2.0	γ = 2.4	γ = 2.8	γ = 3.2	γ = 3.6	Mean	SD
Accuracy	90.20	90.56	91.56	90.23	90.56	90.11	89.99	90.64	89.89	90.41	0.47
Sensitivity	89.65	88.78	89.78	90.12	90.56	88.98	89.15	90.13	91.33	89.83	0.76
Specificity	89.98	90.34	91.23	91.48	91.37	90.49	90.03	89.99	88.95	90.42	0.90
F score	0.88	0.89	0.91	0.91	0.90	0.90	0.91	0.92	0.91	0.90	0.01
Recall	0.88	0.90	0.91	0.91	0.91	0.90	0.89	0.91	0.90	0.90	0.01
ROC area	0.89	0.88	0.89	0.90	.091	0.90	0.91	0.89	0.90	0.91	0.02

4.1. Feature selection

The attribute selection process is significant in improving the analysis of a dataset by eliminating unnecessary features, resulting in reduced memory consumption and increased computational efficiency. In this research, the efficiency of the m-health system is examined by applying it to the entire dataset as well as to three subsets of the dataset created by selecting specific features from the data using the following feature selection methods:

4.1.1. Information gain

Information gain is a measure used in decision to assess the relevance of a feature in a dataset. It is a measure of how much a feature reduces entropy or randomness in the data, and is calculated as the reduction in entropy (uncertainty) of the target variable after a feature is used to split the data into smaller groups. Features with high information gain are considered to be more important or relevant in predicting the target variable. Information gain ranges from 0 to 1, with 0 indicating no reduction in entropy and 1 indicating that the feature perfectly separates the target classes.

We set 0.6 as the threshold value for determining relevant features and removed any irrelevant features as shown in Fig. 4.

4.1.2. Correlation technique

This evaluates the capability of each feature to predict the target class. It allows us to choose the set of features that have a strong relationship with the target class. We have set a threshold of 0.65 for determining significant features and any features below this value were removed based on Fig. 4.

4.1.3. Principal Component Analysis (PCA)

PCA is a statistical technique for dimensionality reduction in which a large set of variables is transformed into a smaller set of uncorrelated variables called principal components. PCA can also be used in feature selection by reducing the number of features in a dataset. In PCA, the features are transformed into a set of linearly uncorrelated principal components. The first principal component contains the most information and each subsequent component contains less and less information. By retaining only the first few principal components, which contain the

Table 5
Statistical results obtained when incorporating SVM with RBF and polynomial kernel for different values of q and γ (Considering features selected using information gain method).

q (RBF)	q = 0.4	q = 0.8	q = 1.2	q = 1.6	q = 2.0	q = 2.4	q = 2.8	q = 3.2	q = 3.6	Mean	S.D
Accuracy (%)	90.23	91.56	92.65	93.56	91.56	92.34	94.65	94.65	92.65	92.65	1.38
Sensitivity (%)	89.99	90.56	91.65	92.36	91.64	92.34	93.02	89.99	90.13	91.29	1.09
Specificity (%)	89.65	90.65	91.35	91.46	91.37	91.89	89.88	90.54	90.49	90.80	0.71
F score	0.88	0.89	0.90	0.87	0.86	0.87	0.89	0.90	0.91	0.88	0.01
Recall	0.90	0.91	0.92	0.89	0.93	0.92	0.90	0.91	0.93	0.91	0.01
ROC area	0.90	0.88	0.90	0.89	0.87	0.92	0.91	0.90	0.93	0.90	0.01
q (polynomial)	q = 0.4	q = 0.8	q = 1.2	q = 1.6	q = 2.0	q = 2.4	q = 2.8	q = 3.2	q = 3.6	Mean	SD
Accuracy (%)	89.99	90.56	93.23	91.56	93.56	91.59	90.53	91.67	91.55	91.58	1.12
Sensitivity (%)	89.66	90.56	91.65	92.32	90.45	90.56	90.68	91.56	90.89	90.92	0.74
Specificity (%)	90.88	90.35	91.56	92.67	93.56	94.65	91.66	93.56	91.32	92.24	1.35
F score	0.89	0.90	0.93	0.91	0.92	0.89	0.88	0.89	0.87	0.89	0.01
Recall	0.90	0.91	0.93	0.89	0.94	0.92	0.91	0.93	0.93	0.91	0.01
ROC area	0.89	0.91	0.90	0.93	0.89	0.88	0.91	0.92	0.93	0.90	0.01
γ (RBF)	γ = 0.4	γ = 0.8	γ = 1.2	γ = 1.6	γ = 2.0	γ = 2.4	γ = 2.8	γ = 3.2	γ = 3.6	Mean	SD
Accuracy	90.65	91.56	89.49	90.44	93.57	91.65	91.89	89.99	91.47	90.35	1.13
Sensitivity	89.99	89.66	90.65	91.34	92.78	90.88	91.66	93.45	91.65	90.44	0.86
Specificity	91.33	92.56	94.65	94.65	90.66	90.54	91.36	92.65	91.66	91.52	1.24
F score	0.90	0.89	0.87	0.90	0.92	0.93	0.91	0.89	0.88	0.90	0.01
Recall	0.89	0.91	0.92	0.90	0.89	0.88	0.90	0.91	0.92	0.90	0.01
ROC area	0.91	0.92	0.91	0.91	0.93	0.89	0.93	0.92	0.91	0.91	0.01
γ (polynomial)	γ = 0.4	γ = 0.8	γ = 1.2	γ = 1.6	γ = 2.0	γ = 2.4	γ = 2.8	γ = 3.2	γ = 3.6	Mean	SD
Accuracy	89.99	90.65	91.36	94.56	91.66	91.54	91.65	92.32	91.56	91.35	1.23
Sensitivity	90.66	91.58	92.33	93.42	91.54	90.65	94.65	91.58	91.56	92.66	1.41
Specificity	89.99	89.54	90.65	92.25	92.35	91.66	91.33	94.12	92.64	90.37	0.86
F score	0.90	0.91	0.91	0.93	0.92	0.89	0.90	0.93	0.94	0.91	0.01
Recall	0.91	0.92	0.9	0.91	0.92	0.91	0.89	0.91	0.92	0.91	0.01
ROC area	0.91	0.92	0.93	0.94	0.92	0.89	0.93	0.91	0.91	0.90	0.01

majority of the information, PCA can be used to effectively reduce the number of features in a dataset.

In the proposed research, PCA is used in combination with other methods, such as the correlation method or information gain, to further improve the feature selection process. By combining PCA with these methods, it is possible to select the most important features and reduce the dimensionality of the dataset in an efficient manner. We chose the principal components that received at least 50% of the ranking. Threshold is set based on the priority-based selection method (i.e., if any feature has a value greater than 0.5, priority will be given to that feature and will be retained in the dataset). Fig. 4 shows that we obtained four new parameters that are a combination of several features.

4.2. Classification performance

We performed a set of experiments where the values of the q exponent and γ parameter are altered in both the RBF and polynomial kernel expression. The performance of accuracy, ROC area, specificity, and sensitivity is assessed across the entire dataset as well as the three subsets derived from the feature selection techniques outlined in Section 4.1. The results are displayed in Tables 4–7.

Table 4 displays the statistical results of the proposed work's performance when all features are selected. The best performance is achieved with an RBF kernel having a q value of 2.4 on the dataset that includes all parameters. With this value, the proposed system has an accuracy of 91.67% in classifying healthy and pathological voices with dysphonia, and a sensitivity of 91.73%.

When using SVM, with a polynomial kernel, the parameter γ represents the kernel coefficient. Higher values of γ can lead to a more complex decision boundary, potentially causing overfitting, while lower values might result in a smoother decision boundary. In the proposed research, it is found that a γ-value of 1.6, detailed in Table 5, produced the optimal performance. This value corresponded to an accuracy rate of 94.56%, a sensitivity of 93.42%, a specificity of 92.25%, and an ROC area of 0.94%. These outcomes are achieved by employing the parameters selected through the information gain method in the SVM utilizing a polynomial kernel.

By incorporating the features selected using the correlation method, the highest accuracy achieved of 95.68%, the sensitivity of 94.89%, specificity of 95.66% and ROC area is 0.91 when the kernel is polynomial and γ-value is 2.8 as shown in Table 6.

Finally, using only parameters selected through PCA, the highest accuracy is obtained using an RBF kernel with a γ-value of 3.2, yielding 99.97%, a sensitivity of 98.64%, specificity of 98.51% and the ROC area is 0.98. The detailed statistical results obtained when only considering features selected using PCA are displayed in Table 7.

The most accurate result (99.97% accuracy) in detecting speech and voice disorders is attained with a polynomial kernel, using all classification parameters and the gamma value of 1. This result is verified through subsequent experiments, where features are selected using two of the three feature selection methods. Specifically, when only features selected by PCA are considered, the highest classification accuracy is achieved using SMO (96.16% and 98.75%, respectively). However, using correlated features, on the other hand, the decision tree algorithm is the most accurate for the classification of a healthy voice and a pathological voice, with an accuracy of 98.97%. Moreover, various machine learning algorithms, such as SVM, DT, NB, RF, and KNN are analyzed to determine the best fit for the system being proposed.

The existing models are implemented in Python 3.7, and the results presented in Table 8 are computed using this implementation. Table 8 displays the results of performance metrics for these algorithms where it can be seen that SVM (with an accuracy of 99.97%) and DT (with an accuracy of 98.96%) are the best options for classifying healthy and pathological voices when the features are selected using PCA. In addition, a thorough comparison is also made between the proposed and existing models. The results represented in Table 8 show that the proposed model classifies healthy and pathological voices with a high degree of accuracy as compared to the existing models. While existing

Table 6
Statistical results obtained when incorporating SVM with RBF and polynomial kernel for different values of q and γ (Considering features selected using Correlation method).

q (RBF)	q = 0.4	q = 0.8	q = 1.2	q = 1.6	q = 2.0	q = 2.4	q = 2.8	q = 3.2	q = 3.6	Mean	S.D
Accuracy (%)	93.12	92.34	91.22	93.45	93.89	91.59	92.59	91.99	92.49	92.36	1.56
Sensitivity (%)	92.34	91.26	92.49	92.10	92.33	92.14	90.14	92.66	92.45	91.56	1.06
Specificity (%)	91.25	92.48	92.67	92.89	92.99	92.15	91.99	91.89	92.98	92.64	1.48
F score	0.90	0.91	0.92	0.91	0.90	0.93	0.92	0.91	0.93	0.91	0.01
Recall	0.91	0.90	0.91	0.92	0.93	0.91	0.92	0.91	0.93	0.92	0.01
ROC area	0.89	0.88	0.90	0.89	0.88	0.88	0.87	0.89	0.90	0.90	0.01
q (polynomial)	q = 0.4	q = 0.8	q = 1.2	q = 1.6	q = 2.0	q = 2.4	q = 2.8	q = 3.2	q = 3.6	92.35	.48
Accuracy (%)	92.11	91.56	92.33	92.89	91.89	91.99	92.49	92.89	92.33	91.86	1.36
Sensitivity (%)	92.31	92.02	91.22	91.56	91.49	91.57	91.66	91.33	91.78	92.67	1.55
Specificity (%)	92.45	91.66	91.48	92.46	91.01	91.65	92.12	92.33	92.45	91.65	1.67
F score	0.91	0.90	0.930	0.92	0.91	0.92	0.90	0.91	0.91	0.91	0.01
Recall	0.91	0.93	0.92	0.91	0.91	0.92	0.90	0.90	0.91	0.90	0.01
ROC area	0.88	0.87	0.89	0.89	0.90	.087	0.88	0.89	0.88	0.92	0.01
γ (RBF)	γ = 0.4	γ = 0.8	γ = 1.2	γ = 1.6	γ = 2.0	γ = 2.4	γ = 2.8	γ = 3.2	γ = 3.6	Mean	SD
Accuracy (%)	91.22	91.32	92.54	93.22	91.45	92.33	93.01	93.00	93.89	92.87	1.74
Sensitivity (%)	92.45	91.64	92.44	93.45	91.56	91.33	92.54	92.36	92.45	91.53	1.08
Specificity (%)	92.15	92.14	93.48	91.00	91.15	91.57	92.88	92.98	92.36	92.57	1.56
F score	0.91	0.90	0.91	0.90	0.92	0.91	0.91	0.92	0.90	0.90	0.01
Recall	0.91	0.92	0.91	0.93	0.92	0.91	0.90	0.90	0.91	0.91	0.01
ROC area	0.88	0.89	0.87	0.88	0.88	0.89	0.87	0.88	0.90	0.90	0.01
γ (polynomial)	γ = 0.4	γ = 0.8	γ = 1.2	γ = 1.6	γ = 2.0	γ = 2.4	γ = 2.8	γ = 3.2	γ = 3.6	Mean	SD
Accuracy (%)	91.23	92.32	91.20	90.45	92.44	91.24	95.65	93.03	91.77	93.64	1.12
Sensitivity (%)	91.23	91.57	91.68	91.44	91.67	92.45	94.89	91.33	92.79	92.45	1.43
Specificity (%)	92.33	92.48	92.48	92.98	93.21	92.11	95.66	91.22	93.94	92.45	1.30
F score	0.91	0.93	0.92	0.90	.091	0.91	0.4	0.93	0.91	0.91	0.01
Recall	0.91	0.92	0.90	0.90	0.91	0.91	0.93	0.90	0.91	0.90	0.01
ROC area	0.88	0.89	0.88	0.89	0.87	0.86	0.91	0.90	0.89	0.91	0.01

Table 7
Statistical results obtained when incorporating SVM with RBF and polynomial kernel for different values of q and γ (Considering features selected using Correlation method).

q (RBF)	q = 0.4	q = 0.8	q = 1.2	q = 1.6	q = 2.0	q = 2.4	q = 2.8	q = 3.2	q = 3.6	Mean	SD
Accuracy (%)	97.56	96.99	97.45	97.89	97.89	97.99	97.86	98.02	97.03	97.63	0.37
Sensitivity (%)	97.56	97.10	98.12	97.06	96.99	96.67	97.55	97.89	97.66	97.51	0.38
Specificity (%)	97.55	97.61	97.49	97.98	96.87	96.88	97.05	97.65	97.45	96.89	0.40
F score	0.97	0.96	0.97	0.96	0.97	0.96	0.98	0.96	0.97	0.96	0.01
Recall	0.97	0.98	0.97	0.96	0.97	0.96	0.98	0.97	0.97	0.97	0.01
ROC area	0.97	0.96	0.97	0.95	0.96	0.95	0.97	0.96	0.96	0.96	0.01
q (polynomial)	q = 0.4	q = 0.8	q = 1.2	q = 1.6	q = 2.0	q = 2.4	q = 2.8	q = 3.2	q = 3.6	Mean	SD
Accuracy (%)	97.65	97.66	97.89	97.46	96.88	96.99	97.64	97.66	97.65	94.43	0.35
Sensitivity (%)	97.66	97.54	97.12	96.84	98.66	97.65	97.16	97.66	97.40	97.54	0.37
Specificity (%)	97.61	97.00	97.13	97.43	97.41	96.49	97.66	97.15	97.54	97.34	0.36
F score	0.97	0.96	0.97	0.96	0.98	0.97	0.96	0.97	0.96	0.96	0.01
Recall	0.97	0.96	0.96	0.97	0.98	0.97	0.96	0.97	0.96	0.97	0.01
ROC area	0.96	0.97	0.96	0.96	0.96	0.97	0.95	0.96	0.97	0.97	0.01
γ (RBF)	γ = 0.4	γ = 0.8	γ = 1.2	γ = 1.6	γ = 2.0	γ = 2.4	γ = 2.8	γ = 3.2	γ = 3.6	Mean	SD
Accuracy (%)	97.66	97.67	97.89	98.66	97.99	98.78	98.66	99.97	97.98	97.24	0.32
Sensitivity (%)	97.66	97.46	97.89	98.00	97.59	97.99	97.6	98.64	97.99	97.62	0.35
Specificity (%)	97.66	97.46	97.49	96.89	98.88	98.46	97.67	98.51	97.78	97.33	0.36
F score	0.98	0.97	0.96	0.97	0.97	0.97	0.97	0.96	0.97	0.96	0.01
Recall	0.97	0.96	0.97	0.97	0.97	0.97	0.96	0.97	0.97	0.97	0.01
ROC area	0.97	0.96	0.97	0.97	0.97	0.96	0.97	0.988	0.97	0.096	0.01
γ (polynomial)	γ = 0.4	γ = 0.8	γ = 1.2	γ = 1.6	γ = 2.0	γ = 2.4	γ = 2.8	γ = 3.2	γ = 3.6	Mean	SD
Accuracy (%)	97.65	97.64	96.99	97.56	97.66	96.98	96.66	97.89	97.89	97.37	0.35
Sensitivity (%)	97.66	97.49	98.66	96.66	97.66	97.89	98.84	97.88	97.65	97.16	0.36
Specificity (%)	97.66	97.89	97.88	97.66	967.46	97.65	97.65	97.66	97.79	97.37	0.35
F score	0.96	0.97	0.96	0.97	0.96	0.97	0.98	0.97	0.97	0.97	0.01
Recall	0.96	0.97	0.96	0.97	0.97	0.98	00.96	0.97	0.97	0.96	0.01
ROC area	0.97	0.96	0.97	0.97	0.96	0.97	0.96	0.97	0.96	0.97	0.01

models for voice order detection typically offer accuracy of around 97%, the proposed model surpasses this benchmark, achieving over 97% accuracy. This suggests that prioritizing the proposed model is ideal for ensuring accurate output results. The achieved test accuracies demonstrate that the proposed model is capable of identifying voice disorders in unknown speakers as well.

Fig. 5 displays the optimal results achieved by the proposed model utilizing SVM with features selected through PCA. Fig. 5(a) shows the model's confusion matrix without K-fold analysis. Additionally, Fig. 5(b) and (c) represent the confusion matrices obtained through 5-fold and 10-fold analyses, respectively.

Recent studies have demonstrated commendable accuracy in machine learning models. However, upon comparison with our proposed

Table 8
Performance comparison of several machine learning algorithms.

When considering all features					
Metrics	SVM	DT	NB	RF	KNN
Accuracy (%)	96.56	96.78	97.66	97.56	97.33
Sensitivity (%)	97.65	98.56	97.88	97.44	97.60
Specificity (%)	97.56	98.61	97.26	97.65	97.55
F score	0.97	0.98	0.97	0.96	0.97
Recall	0.96	0.97	0.96	0.97	0.97
ROC area	0.98	.097	0.96	0.97	0.98
When considering the features selected using information gain method					
Accuracy (%)	97.65	98.66	97.64	96.44	97.31
Sensitivity (%)	96.67	96.88	97.45	97.61	97.66
Specificity (%)	97.66	97.41	98.31	96.66	97.86
F score	0.97	0.96	0.97	0.9	0.97
Recall	0.97	0.96	0.96	0.97	0.98
ROC area	0.960	.97	0.97	0.98	0.97
When considering the features selected using correlation method					
Accuracy (%)	97.66	97.45	96.99	96.87	97.16
Sensitivity (%)	97.64	97.55	97.48	96.88	97.90
Specificity (%)	97.66	97.48	97.61	97.21	98.03
F score	0.97	0.96	0.97	0.98	0.96
Recall	0.97	.096	0.98	0.97	0.96
ROC area	0.97	0.97	0.96	0.98	0.97
When considering the features selected using PCA					
Accuracy (%)	99.897	98.86	98.67	98.88	98.91
Sensitivity (%)	98.64	98.66	97.66	98.64	97.66
Specificity (%)	98.51	98.12	97.61	98.31	97.66
F score	0.99	0.98	0.96	0.97	0.98
Recall	0.99	0.96	0.96	0.98	0.97
ROC area	0.98	0.98	0.96	0.97	0.97
Comparison with the existing models					
Ref	Schlegel et al. (2020)	Al-Hussain et al. (2022)	Cesari et al. (2018)	Alhussein and Muhammad (2018)	Darouiche et al. (2022)
Accuracy (%)	98.65	97.55	97.84	97.65	97.88
Sensitivity (%)	97.56	97.65	96.99	97.45	97.60
Specificity (%)	97.64	97.12	98.36	96.45	97.45
F score	0.96	0.97	0.97	0.97	0.96
Recall	0.98	0.96	0.97	0.97	0.96
ROC area	0.97	0.96	0.98	0.97	0.96

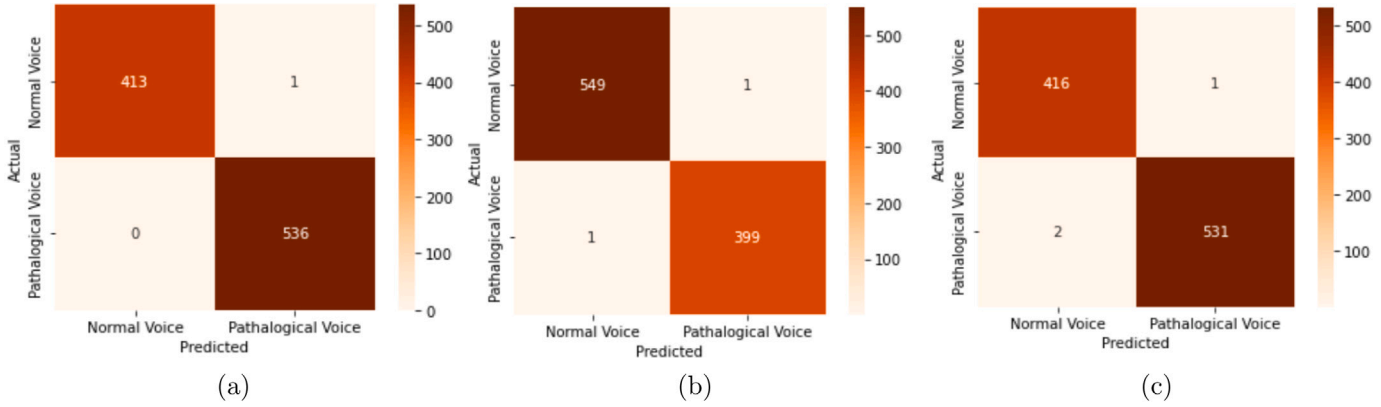


Fig. 5. Confusion matrices for the proposed model: (a) When SVM is applied with No K-folds, (b) When SVM is applied with 5 K-folds, (c) When SVM is applied with 10 K-folds.

model, their accuracy diminishes, highlighting the superior performance and reliability of our approach for real-time applications. Furthermore, existing studies (Schlegel et al., 2020; Cesari et al., 2018) utilized Phonovibrogram (PVG)-based features and cepstral features rather than the characteristic voice signal features employed in our model. This distinction shows the uniqueness of our proposed model compared to existing ones, offering a novel approach in this domain.

4.2.1. K-fold analysis

The proposed study extends the evaluation of the proposed learning model by employing K-fold validation. This technique involves

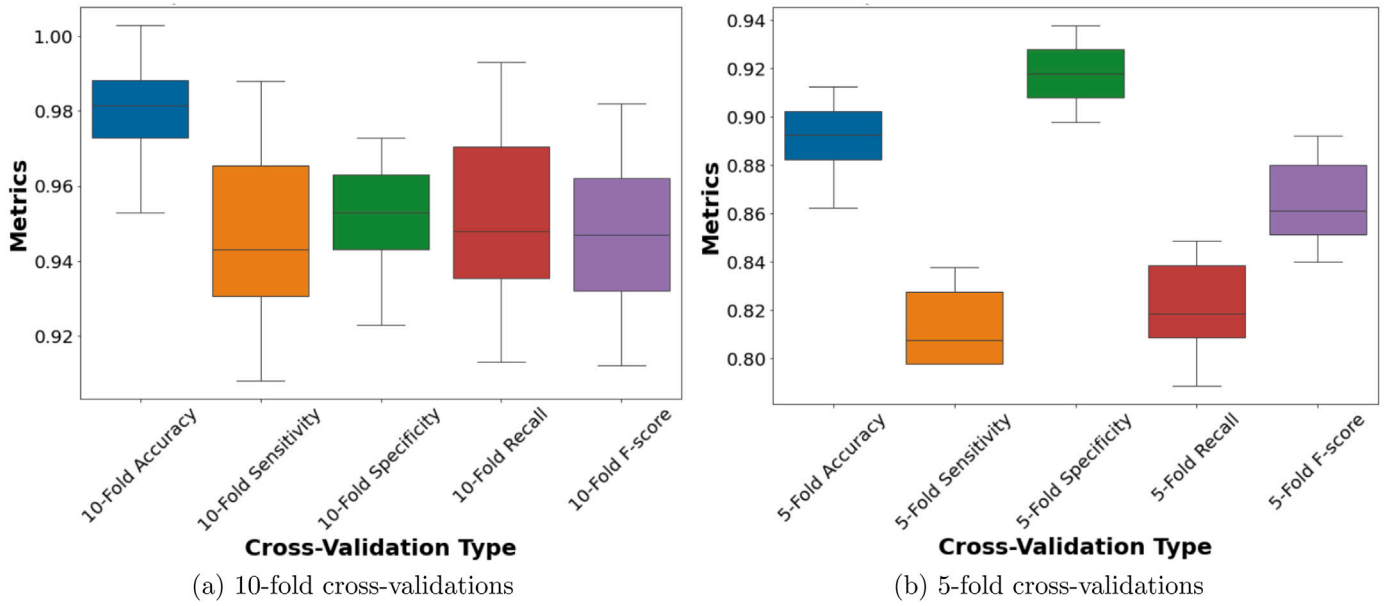


Fig. 6. Boxplots for the 10-fold and 5-fold cross-validations.

Table 9
K-fold analysis (When considering the features selected using PCA).

Parameters →	SVM	DT	NB	RF	KNN
Accuracy					
K-folds↓					
K = 5	99.78	98.86	98.66	98.76	98.79
K = 10	99.68	98.64	99.66	98.83	98.71
Sensitivity (%)					
K = 5	0.95	0.32	0.91	0.92	
K = 10	0.96	0.34	0.83	0.83	1.00
Specificity (%)					
K = 5	0.99	1.00	0.90	0.90	0.78
K = 10	0.98	1.00	0.80	0.80	0.60
F score (%)					
K = 5	0.98	0.33	0.94	0.95	0.99
K = 10	0.97	0.33	0.96	0.96	1.00
Recall (%)					
K = 5	1.00	1.00	0.93	0.93	0.79
K = 10	0.99	1.00	0.95	0.95	0.81
ROC area					
K = 5	97	50	94	95	89
K = 10	95	50	95	96	90

assessing the model's performance by rotating each dataset sample as a testing point while utilizing the rest for training. The proposed model is evaluated using 5-fold and 10-fold cross-validation tests. Table 9 shows the outcomes of these K-fold analyses, revealing the model's exceptional accuracy, which is more than 98%. This robust performance solidifies the model's effectiveness as demonstrated in the study. Moreover, boxplots are generated as shown in Fig. 6 to display the top-performing accuracy, sensitivity, specificity, recall, and F-score achieved by the proposed method during both the 10-fold and 5-fold cross-validations.

5. Conclusions, limitation and future work

Recently, the use of mobile apps and multimedia services in the healthcare industry has grown rapidly. These applications give people the convenience of accessing important medical information and data from anywhere at any time. This is especially helpful in monitoring, identifying and detecting voice pathologies, such as dysphonia, which is a common voice disorder often ignored but affects many people.

There has been a lot of attention given to researching mobile systems that can automatically identify voice disorders. This is because these systems are objective and non-invasive. Machine learning algorithms can aid in exploring new methods of speech and voice signal processing and be easily integrated into m-health solutions. This study compares the effectiveness of various voice pathology identification methods using machine learning techniques, such as Decision Tree, Support Vector Machine, Logistic Model Tree and Naïve Bayes. The proposed study also focuses on determining the best voice signal features to use by comparing different classifiers. All the analyses are done using a large dataset (CMED) of 8158 voice samples.

In the proposed study, several experimental results and analyses are conducted on the entire dataset and three subsets using three different feature selection methods. The results showed that the Support Vector Machine algorithm provided the highest accuracy in detecting voice pathologies, with an accuracy rate of around 99.97% when the features are selected using PCA.

The speech samples in the dataset are not recorded using any particular microphone or acoustic transducer. However, the transducers can have a significant impact on voice and speech. Therefore, determining the precise scale or level of such vocal variations that are only based on acoustic transducers would require further analysis and context-specific information. The proposed model in this research only works on speech samples not recorded by any acoustic transducer. This limitation of the model will be addressed and improved in the future.

Furthermore, existing studies in the literature have reported lower accuracy levels, primarily due to the limited and often inaccessible datasets used. Our future focus is to enhance the classification outcomes within telemedicine. We aim to integrate hybrid classifiers and deep learning algorithms into a mobile health platform, enabling the detection of voice disorders and facilitating patient monitoring and treatment remotely. Additionally, we plan to employ paraconsistent feature engineering techniques, including feature fusion, adaptive model learning, ensemble methods, fuzzy logic, and probabilistic approaches, to augment our proposed model. Exploring the incorporation of signal mass and the Enhanced Teager Energy Operator (ETEO) as features in voice analysis algorithms will also be a consideration, further bolstering the capabilities of our model in telemedicine applications.

CRediT authorship contribution statement

Mujeeb Ur Rehman: Writing – original draft. **Arslan Shafique:** Originator of the idea and lead in its implementation. **Qurat-Ul-Ain Azhar:** Conducted analysis, Compiled the results. **Sajjad Shaukat Jamal:** Provided essential funding, Resources for the project. **Aminu Bello Usman:** Contributed by conducting a thorough review of the work.

Declaration of competing interest

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

Data availability

The data that has been used is confidential.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University, for funding this work through large group Research Project, under Grant R. G. P. 2/238/44.

References

Al-Dhief, F.T., Baki, M.M., Latiff, N.M.A., Malik, N.N.N.A., Salim, N.S., Albader, M.A.A., Mahyuddin, N.M., Mohammed, M.A., 2021. Voice pathology detection and classification by adopting online sequential extreme learning machine. IEEE Access 9, 77293–77306.

Al-Dhief, F.T., Latiff, N.M.A., Malik, N.N.N.A., Baki, M.M., Sabri, N., Albadr, M.A.A., 2022. Dysphonia detection based on voice signals using naive Bayes classifier. In: 2022 IEEE 6th International Symposium on Telecommunication Technologies. ISTT, IEEE, pp. 56–61.

AL-Dhief, F.T., Latiff, N.M.A., Malik, N.N.N.A., Sabri, N., Baki, M.M., Albadr, M.A.A., Abbas, A.F., Hussein, Y.M., Mohammed, M.A., 2020. Voice pathology detection using machine learning technique. In: 2020 IEEE 5th International Symposium on Telecommunication Technologies. ISTT, IEEE, pp. 99–104.

Al-Dhief, F.T., Latiff, N.M.A., Malik, N.N.N.A., Salim, N.S., Baki, M.M., Albadr, M.A.A., Mohammed, M.A., 2020. A survey of voice pathology surveillance systems based on internet of things and machine learning algorithms. IEEE Access 8, 64514–64533.

Al-Hussain, G., Shuweihdi, F., Alali, H., Househ, M., Abd-Alrazaq, A., 2022. The effectiveness of supervised machine learning in screening and diagnosing voice disorders: Systematic review and meta-analysis. J. Med. Internet Res. 24 (10), e38472.

Al Mojaly, M., Muhammad, G., Alsulaiman, M., 2014. Detection and classification of voice pathology using feature selection. In: 2014 IEEE/ACS 11th International Conference on Computer Systems and Applications. AICCSA, IEEE, pp. 571–577.

Alaiad, A., Alsharo, M., Alnsour, Y., 2019. The determinants of m-health adoption in developing countries: an empirical investigation. Appl. Clin. Inform. 10 (05), 820–840.

Albadr, M.A.A., Ayob, M., Tiun, S., AL-Dhief, F.T., Arram, A., Khalaf, S., 2023. Breast cancer diagnosis using the fast learning network algorithm. Front. Oncol. 13, 1150840.

Albadr, M.A.A., Ayob, M., Tiun, S., Al-Dhief, F.T., Hasan, M.K., 2022a. Gray wolf optimization-extreme learning machine approach for diabetic retinopathy detection. Front. Public Health 10, 925901.

Albadr, M.A.A., Tiun, S., Ayob, M., Al-Dhief, F.T., 2022b. Particle swarm optimization-based extreme learning machine for covid-19 detection. Cogn. Comput. 1–16.

Albadr, M.A.A., Tiun, S., Ayob, M., Al-Dhief, F.T., Abdali, T.-A.N., Abbas, A.F., 2021. Extreme learning machine for automatic language identification utilizing emotion speech data. In: 2021 International Conference on Electrical, Communication, and Computer Engineering. ICECCE, IEEE, pp. 1–6.

Albadr, M.A.A., Tiun, S., Ayob, M., AL-Dhief, F.T., Omar, K., Maen, M.K., 2022c. Speech emotion recognition using optimized genetic algorithm-extreme learning machine. Multimedia Tools Appl. 81 (17), 23963–23989.

Alhussein, M., Muhammad, G., 2018. Voice pathology detection using deep learning on mobile healthcare framework. IEEE Access 6, 41034–41041.

Alhussein, M., Muhammad, G., 2019. Automatic voice pathology monitoring using parallel deep models for smart healthcare. IEEE Access 7, 46474–46479.

Ali, Z., Muhammad, G., Alhamid, M.F., 2017. An automatic health monitoring system for patients suffering from voice complications in smart cities. IEEE Access 5, 3900–3908.

Amara, F., Fezari, M., Bourouba, H., 2016. An improved GMM-SVM system based on distance metric for voice pathology detection. Appl. Math. 10 (3), 1061–1070.

Angerstein, W., Baracca, G., Dejonckere, P., Echternach, M., Eysholdt, U., Fussi, F., Geneid, A., Hacki, T., Karmelita-Katulska, K., Haubrich, R., et al., 2019. Diagnosis and differential diagnosis of voice disorders. In: Phoniatrics I: Fundamentals–Voice Disorders–Disorders of Language and Hearing Development. Springer, pp. 349–430.

Arora, R., 2012. Comparative analysis of classification algorithms on different datasets using WEKA. Int. J. Comput. Appl. 54 (13).

Barreira, R.R., Ling, L.L., 2020. Kullback–Leibler divergence and sample skewness for pathological voice quality assessment. Biomed. Signal Process. Control 57, 101697.

Bhowmik, S., Hasan, M., Hakim, M.A., 2022. A dimensionality reduction based efficient multiple voice disease recognition scheme using mel-frequency cepstral coefficients and K-nearest neighbors algorithm. In: Proceedings of the International Conference on Big Data, IoT, and Machine Learning: BIM 2021. Springer, pp. 301–313.

Boyanov, B., Hadjitodorov, S., 1997. Acoustic analysis of pathological voices. A voice analysis system for the screening of laryngeal diseases. IEEE Eng. Med. Biol. Mag. 16 (4), 74–82.

Brockmann, M., Storck, C., Carding, P.N., Drinnan, M.J., 2008. Voice loudness and gender effects on jitter and shimmer in healthy adults.

Calvo, R.A., D'Mello, S., 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Trans. Affect. Comput. 1 (1), 18–37.

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

Chaiani, M., Selouani, S.A., Boudraa, M., Yakoub, M.S., 2022. Voice disorder classification using speech enhancement and deep learning models. Biocybern. Biomed. Eng. 42 (2), 463–480.

Chen, Z., Zhu, P., Qiu, W., Guo, J., Li, Y., 2023. Deep learning in automatic detection of dysphonia: Comparing acoustic features and developing a generalizable framework. Int. J. Lang. Commun. Disord. 58 (2), 279–294.

Cordeiro, H., Fonseca, J., Guimarães, I., Meneses, C., 2015. Voice pathologies identification speech signals, features and classifiers evaluation. In: 2015 Signal Processing: Algorithms, Architectures, Arrangements, and Applications. SPA, IEEE, pp. 81–86.

Darouiche, M.S., El Moubtahij, H., Yakhlef, M.B., Tazi, E.B., 2022. An automatic voice disorder detection system based on extreme gradient boosting classifier. In: 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology. IRASET, IEEE, pp. 1–5.

Dias, D., Paulo Silva Cunha, J., 2018. Wearable health devices—vital sign monitoring, systems and technologies. Sensors 18 (8), 2414.

El Emary, I., Fezari, M., Amara, F., 2014. Towards developing a voice pathologies detection system. J. Commun. Technol. Electron. 59, 1280–1288.

Ezugwu, A.E., Ikotun, A.M., Oyelade, O.O., Abualigah, L., Agushaka, J.O., Eke, C.I., Akinyelu, A.A., 2022. A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. Eng. Appl. Artif. Intell. 110, 104743.

Fang, S.-H., Tsao, Y., Hsiao, M.-J., Chen, J.-Y., Lai, Y.-H., Lin, F.-C., Wang, C.-T., 2019. Detection of pathological voice using cepstrum vectors: A deep learning approach. J. Voice 33 (5), 634–641.

Fonseca, E.S., Guido, R.C., Junior, S.B., Dezani, H., Gati, R.R., Pereira, D.C.M., 2020. Acoustic investigation of speech pathologies based on the discriminative paraconsistent machine (DPM). Biomed. Signal Process. Control 55, 101615.

Fonseca, E.S., Guido, R.C., Scalassara, P.R., Maciel, C.D., Pereira, J.C., 2007. Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders. Comput. Biol. Med. 37 (4), 571–578.

García, L., Tomás, J., Parra, L., Lloret, J., 2019. An m-health application for cerebral stroke detection and monitoring using cloud services. Int. J. Inf. Manage. 45, 319–327.

Geng, L., Liang, Y., Shan, H., Xiao, Z., Wang, W., Wei, M., 2022. Pathological voice detection and classification based on multimodal transmission network. J. Voice.

Gidaye, G., Nirmal, J., Ezzine, K., Frikha, M., 2022. Unified wavelet-based framework for evaluation of voice impairment. Int. J. Speech Technol. 25 (2), 527–548.

Godino-Llorente, J.I., Gómez-Vilda, P., Sáenz-Lechón, N., Blanco-Velasco, M., Cruz-Roldán, F., Ferrer-Ballester, M.A., 2005. Support vector machines applied to the detection of voice disorders. In: Nonlinear Analyses and Algorithms for Speech Processing: International Conference on Non-Linear Speech Processing, NOLISP 2005, Barcelona, Spain, April 19-22, 2005, Revised Selected Papers. Springer, pp. 219–230.

Harar, P., Alonso-Hernandezy, J.B., Mekyska, J., Galaz, Z., Burget, R., Smekal, Z., 2017. Voice pathology detection using deep learning: a preliminary study. In: 2017 International Conference and Workshop on Bioinspired Intelligence. IWOBI, IEEE, pp. 1–4.

Hossain, M.S., 2016. Patient state recognition system for healthcare using speech and facial expressions. J. Med. Syst. 40, 1–8.

Idrisoglu, A., Dallora, A.L., Anderberg, P., Berglund, J.S., 2023. Applied machine learning techniques to diagnose voice-affecting conditions and disorders: systematic literature review. J. Med. Internet Res. 25, e46105.

Jothilakshmi, S., 2014. Automatic system to detect the type of voice pathology. Appl. Soft Comput. 21, 244–249.

Kelly, A.C., Gobl, C., 2011. A comparison of mel-frequency cepstral coefficient (MFCC) calculation techniques. J. Comput. 3 (10), 62–66.

Kim, H., Jeon, J., Han, Y.J., Joo, Y., Lee, J., Lee, S., Im, S., 2020. Convolutional neural network classifies pathological voice change in laryngeal cancer with high accuracy. J. Clin. Med. 9 (11), 3415.

Kirmayer, L., Simpson, C., Cargo, M., 2003. Healing traditions: Culture, community and mental health promotion with Canadian Aboriginal peoples. Australas. Psychiatry 11 (sup1), S15–S23.

Kwon, S., 2019. A CNN-assisted enhanced audio signal processing for speech emotion recognition. Sensors 20 (1), 183.

Li, X., Tao, J., Johnson, M.T., Soltis, J., Savage, A., Leong, K.M., Newman, J.D., 2007. Stress and emotion classification using jitter and shimmer features. In: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07. Vol. 4, IEEE, pp. IV–1081.

Mamun, M., Mahmud, M.I., Hossain, M.I., Islam, A.M., Ahammed, M.S., Uddin, M.M., 2022. Vocal feature guided detection of parkinson's disease using machine learning algorithms. In: 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference. UEMCON, IEEE, pp. 0566–0572.

Mamyrbayev, O., Mekebayev, N., Turdalyuly, M., Oshanova, N., Medeni, T.I., Yessentay, A., 2019. Voice identification using classification algorithms. Intell. Syst. Comput..

Mesallam, T.A., Farahat, M., Malki, K.H., Alsulaiman, M., Ali, Z., Al-Nasheri, A., Muhammad, G., et al., 2017. Development of the arabic voice pathology database and its evaluation by using speech features and machine learning algorithms. J. Healthc. Eng. 2017.

Mittal, V., Sharma, R., 2021. Deep learning approach for voice pathology detection and classification. Int. J. Healthc. Inf. Syst. Inform. (IJHISI) 16 (4), 1–30.

Mohammed, M.A., Abdulkareem, K.H., Mostafa, S.A., Khanapi Abd Ghani, M., Maashi, M.S., Garcia-Zapirain, B., Oleagordia, I., Alhakami, H., Al-Dhief, F.T., 2020. Voice pathology detection and classification using convolutional neural network model. Appl. Sci. 10 (11), 3723.

Mohammed, H., Omeroglu, A.N., Polat, M., Oral, E.A., Ozbek, I.Y., 2021. Voice pathology classification using machine learning. In: International Conference on Applied Science and Engineering, ISASE. pp. 354–358.

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

Myles, A.J., Feudale, R.N., Liu, Y., Woody, N.A., Brown, S.D., 2004. An introduction to decision tree modeling. J. Chemom. 18 (6), 275–285.

Nicastri, M., Chiarella, G., Gallo, L., Catalano, M., Cassandro, E., et al., 2004. Multidimensional Voice Program (MDVP) and amplitude variation parameters in euphonic adult subjects. Normative study. Acta Otorhinolaryngol. Ital. 24 (6), 337–341.

Paniagua, M.S., Pérez, C.J., Calle-Alonso, F., Salazar, C., 2020. An acoustic-signal-based preventive program for university lecturers' vocal health. J. Voice 34 (1), 88–99.

Philip, N.Y., Rodrigues, J.J., Wang, H., Fong, S.J., Chen, J., 2021. Internet of Things for in-home health monitoring systems: Current advances, challenges and future directions. IEEE J. Sel. Areas Commun. 39 (2), 300–310.

Powell, M.E., Rodriguez Cancio, M., Young, D., Nock, W., Abdelmessih, B., Zeller, A., Perez Morales, I., Zhang, P., Garrett, C.G., Schmidt, D., et al., 2019. Decoding phonation with artificial intelligence (DeP AI): proof of concept. Laryngoscope Investig. Otolaryngol. 4 (3), 328–334.

Reddy, E.M.K., Gurrala, A., Hasitha, V.B., Kumar, K.V.R., 2022. Introduction to naive Bayes and a review on its subtypes with applications. In: Bayesian Reason. Gaussian Process. Mach. Learn. Appl. pp. 1–14.

Reid, J., Parmar, P., Lund, T., Aalto, D.K., Jeffery, C.C., 2022. Development of a machine-learning based voice disorder screening tool. Am. J. Otolaryngol. 43 (2), 103327.

Ritchings, R., McGillion, M., Moore, C.J., 2002. Pathological voice quality assessment using artificial neural networks. Med. Eng. Phys. 24 (7–8), 561–564.

Schlegel, P., Kniesburges, S., Dürr, S., Schützenberger, A., Döllinger, M., 2020. Machine learning based identification of relevant parameters for functional voice disorders derived from endoscopic high-speed recordings. Sci. Rep. 10 (1), 10517.

Seedat, N., Aharonson, V., Hamzany, Y., 2020. Automated and interpretable m-health discrimination of vocal cord pathology enabled by machine learning. In: 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering. CSDE, IEEE, pp. 1–6.

Selvakumari, N.S., Radha, V., 2017. A voice activity detector using SVM and Naïve Bayes classification algorithm. In: 2017 International Conference on Signal Processing and Communication. ICSPC, IEEE, pp. 1–6.

Shafique, A., Mehmood, A., Elhadef, M., 2021. Detecting signal spoofing attack in uavs using machine learning models. IEEE Access 9, 93803–93815.

Sharma, S.K., Al-Wanain, M.I., Alowaidi, M., Alsaghier, H., 2022. Mobile healthcare (m-Health) based on artificial intelligence in healthcare 4.0. Expert Syst. e13025.

Souissi, N., Cherif, A., 2015. Dimensionality reduction for voice disorders identification system based on mel frequency cepstral coefficients and support vector machine. In: 2015 7th International Conference on Modelling, Identification and Control. ICMIC, IEEE, pp. 1–6.

Souissi, N., Cherif, A., 2016. Artificial neural networks and support vector machine for voice disorders identification. Int. J. Adv. Comput. Sci. Appl. 7 (5).

Spadaro, B., Martin-Key, N.A., Funnell, E., Bahn, S., et al., 2022. mHealth solutions for perinatal mental health: Scoping review and appraisal following the mHealth index and navigation database framework. JMIR mHealth uHealth 10 (1), e30724.

Srivastava, R., Shree, R., Shukla, A.K., Pandey, R.P., Shukla, V., Pandey, D., 2022. A feature based classification and analysis of hidden Markov model in speech recognition. In: Cyber Intelligence and Information Retrieval: Proceedings of CIIR 2021. Springer, pp. 365–379.

Subramaniam, S., Majumder, S., Faisal, A.I., Deen, M.J., 2022. Insole-based systems for health monitoring: Current solutions and research challenges. Sensors 22 (2), 438.

Upadhya, S.S., Cheeran, A., 2018. Discriminating Parkinson and healthy people using phonation and cepstral features of speech. Procedia Comput. Sci. 143, 197–202.

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

Vernero, I., Schindler, O., 2012. Storia Della Logopedia, vol. 22, Springer Science & Business Media.

Vizza, P., Tradigo, G., Mirarchi, D., Bossio, R.B., Lombardo, N., Arabia, G., Quattrone, A., Veltri, P., 2019. Methodologies of speech analysis for neurodegenerative diseases evaluation. Int. J. Med. Inform. 122, 45–54.

Wang, J., Jo, C., 2007. Vocal folds disorder detection using pattern recognition methods. In: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 3253–3256.

Xu, M., Zhang, F., Khan, S.U., 2020. Improve accuracy of speech emotion recognition with attention head fusion. In: 2020 10th Annual Computing and Communication Workshop and Conference. CCWC, IEEE, pp. 1058–1064.

Yang, S., Zheng, F., Luo, X., Cai, S., Wu, Y., Liu, K., Wu, M., Chen, J., Krishnan, S., 2014. Effective dysphonia detection using feature dimension reduction and kernel density estimation for patients with Parkinson's disease. PLoS One 9 (2), e88825.

Zhou, C., Wu, Y., Fan, Z., Zhang, X., Wu, D., Tao, Z., 2022. Gammatone spectral latitude features extraction for pathological voice detection and classification. Appl. Acoust. 185, 108417.