www.czasopisma.pan.pl PAN www.journals.pan.pl

ARCHIVES OF ACOUSTICS Vol. 48, No. 1, pp. 13–24 (2023), doi: 10.24425/aoa.2022.142905

Research Paper

Comparative Analysis of Classifiers for the Assessment of Respiratory Disorders Using Speech Parameters

Poonam SHRIVASTAVA[1], Neeta TRIPATHI[1], Bikesh Kumar SINGH[2], Bhupesh Kumar DEWANGAN[3]*

[1] *Department of Electronics and Telecommunication, SSTC*
Bhilai, India

[2] *Department of Biomedical Engineering, National Institute of Technology*
Raipur, India

[3] *Department of Computer Science and Engineering, School of Engineering, OP Jindal University*
Raigarh, India

*Corresponding Author e-mail: bhupesh.dewangan@gmail.com

(*received September 9, 2021; accepted August 31, 2022*)

Non-invasive techniques for the assessment of respiratory disorders have gained increased importance in recent years due to the complexity of conventional methods. In the assessment of respiratory disorders, machine learning may play a very essential role. Respiratory disorders lead to variation in the production of speech as both go hand in hand. Thus, speech analysis can be a useful means for the pre-diagnosis of respiratory disorders. This article aims to develop a machine learning approach to differentiate healthy speech from speech corresponding to different respiratory disorders (affected). Thus, in the present work, a set of 15 relevant and efficient features were extracted from acquired data, and classification was done using different classifiers for healthy and affected speech. To assess the performance of different classifiers, accuracy, specificity (Sp), sensitivity (Se), and area under the receiver operating characteristic curve (AUC) was used by applying both multi-fold cross-validation methods (5-fold and 10-fold) and the holdout method. Out of the studied classifiers, decision tree, support vector machine (SVM), and k-nearest neighbor (KNN) were found more appropriate in providing correct assessment clinically while considering 15 features as well as three significant features (Se > 89%, Sp > 89%, AUC > 82%, and accuracy > 99%). The conclusion was that the proposed classifiers may provide an aid in the simple assessment of respiratory disorders utilising speech parameters with high efficiency. In the future, the proposed approach can be evaluated for the detection of specific respiratory disorders such as asthma, COPD, etc.

Keywords: healthy speech; affected speech; machine learning; classification techniques; respiratory disorders; speech analysis.

1. Introduction

The sound produced by humans to express language orally is called speech. As the respiratory system is the power source, thus, for the production of speech, to produce vibration in the vocal cord, sufficient airflow is required (DOGAN *et al.*, 2007). Different breathing patterns depend on the purpose and nature of speech production. As speaking occurs only during exhalation, to increase the time available for speech production, breath out is slower and breath in is fast. Any kind of airway inflammation can affect the sound of voice quality. Disorders of the respiratory system may affect any of the structures and organs which have to do with breathing (MOHAMED, EL MAGHRABY, 2014). Human speech analysis is a wide research area that helps in medical condition diagnosis affecting the speech parameters (DIXIT *et al.*, 2014). As per the Global Initiative for Chronic Obstructive Lung Diseases (GOLD), respiratory diseases are affecting 400 million worldwide (Halpin *et al.*, 2020). To treat respiratory disorders promptly and appropriately, correct diagnosis is essential. Initially, diagnosis involves auscultation, i.e., the use of a stethoscope for examining lung sounds. Analog filtering and sound amplification is the basic requirement of standard stethoscopes to be interpreted by trained professionals. Then the pulmonary function test (PFT) which measures lungs volume and capacity, airflow rates, and gas exchange is performed. PFT can be done by two methods: spirometry or plethysmography. Along with these, chest X-rays and CT scans are performed. Still, misdiagnosis, under-diagnosis, and delayed diagnosis may occur in the treatment. The reason may involve the expertise needed for performing PFT and auscultation, overlapping among the disease, and the complexity of the disease. The prediction of the disease in the initial stage is very important in the medical field, as death may occur if a proper treatment got delayed.

Thus, we proposed a computerised analysis of speech signals for normal individuals and patients affected by different respiratory disorders.

1.1. Literature review

A comparative study between parametric and non-parametric methods, involving the mathematical transformations in the analysis of speech for the detection of disease has been demonstrated in (Sonu, Sharma, 2012). Mel-frequency cepstral coefficient (MFCC) for feature extraction and dynamic time warping (DTW) for feature matching were used. Although it was a time-consuming process, voice signal could be an alternative approach for respiratory disorder analysis.

A comparison of the performances of various classifiers such as Gaussian mixture model (GMM), multilayer perceptron (MLP) neural networks, support vector machine (SVM), and hierarchical fuzzy signature (HFS) along with the usage of a hybrid classifier, which also reduced the dimensionality, was reported in (Alghowinem *et al.*, 2013). It was observed that the best performance was given by using SVM with GMM as the hybrid classifier. Out of the three fusion methods, it was observed that while associating with HFS, MLP, and GMM, the performance of score fusion was better, while for SVM, the performance of decision fusion was the best. Feature fusion resulted in very poor performance as compared to other methods.

An acoustic analysis for asthmatic and normal persons in which jitter, shimmer, noise to harmonic ratio (NHR), and harmonic to noise ratio (HNR) showed significant variation was reported in (Teixeira, Fernandes, 2014). Jitter, shimmer, and HNR values for males and females were recorded. The result obtained on vowel comparison was found to have no difference between jitter values but there was a difference for shimmer and HNR values.

For the diagnosis of chronic diseases, different algorithms of classification have been applied to the database of diseases and the results are very promising. Still, a novel classification technique is needed. The different methods of acoustic feature extraction and classification that can help in detecting the disease in the prior stage are to be developed so that the process of diagnosis can be simplified.

The different methods of acoustic feature extraction and classification that can help in detecting the disease in the prior stage leading to the discrimination between the voice of healthy and unhealthy persons were discussed in (Saloni *et al.*, 2014). Digital signal processing (DSP) techniques were used for feature extraction whereas vector quantization (VQ), DTW, SVM, GMM, and artificial neural network (ANN) were used for feature classification. It was observed that different classification techniques may not be compared directly due to being measured on a different database.

An automatic disease diagnosis system that was adaptive based on SVM was developed in (Gürbüz, Kiliç, 2014). For the detection of disease in a better way, a new kind of SVM, "Adaptive SVM" has been introduced, showing 100% correct classification rates. The result showed that the proposed method demonstrated a higher success rate than an adaptive method as compared to non-adaptive methods. The method was not disease-specific and as practical as it separates the bias parameters space into subfragments.

To determine the level of asthma, a numerical formula was demonstrated in (Walia, Sharma, 2016). Voice parameters like jitter, shimmer, fundamental frequency, and maximum phonation time were used for generating the formula. On analysis, it was found that the jitter value was low for healthy and high for asthmatic patients while maximum phonation time was vice versa.

The myAirCoach decision support systems design aspects were proposed in (Kocsis *et al.*, 2017) with the focus on the analysis of three machine learning approaches (SVM, random forests, AdaBoot) as support tools. In comparison with SVM and AdaBoot, the random forests algorithm shows better accuracy.

SVM, Naïve Bayes, decision tree, and ANN are considered to be the most widely used classifiers for chronic disease prediction, but Jain and Singh (2018) put focus on adaptive and parallel classification systems that enhance the rate of success and reduce the time taken in making the decision. In this proposed method only for feature selection, the filter method was found to be more efficient. However, by applying hybrid approaches to disease databases, redundant, noisy, and insignificant features may be reduced.

After applying machine learning to self-management asthma, it was found that both Naïve Bayes and logistic regression-based classifiers provided the highest accuracy (AUC > 0.87) in (Tsang *et al.*, 2020). Asthma Mobile Health Study (AMHS) dataset was used. Several prevailing machine learning classifiers, both probabilistic and deterministic models and linear and non-linear were used. Along with AUC, geometric mean accuracy (GMA) was employed due to the skewed nature of the data.

A feasible method of disease detection using analysis of voice was proposed in (Gore *et al.*, 2020). For feature extraction, MFFC and feature matching DTW were applied. Various voice analyses were presented and verified to track characteristics variation in patients' voices.

For the passive assessment of pulmonary functions, two algorithms were proposed in (Chun *et al.*, 2020). One of them was used for distinguishing between healthy and affected with pulmonary disease and the other one to estimate the FEV1/FVC (Forced Expiratory Volume to Forced Vital Capacity) ratio using speech features. Data sets from the research study and in-clinic study were used to develop and validate the algorithms. It was observed that the classification accuracy was obtained to be 73.7% while the F1 score was 84.5%. Also, a mean absolute error of 8.6% was observed with FEV1/FVC ratio in regression analysis.

Even though so much work has already been done in the speech analysis area and respiratory disorders but still, less work has been done for combining both. In the speech area, recognition and emotional patterns have been considered while the respiratory function is generally assessment done using lung sounds. So, this paper deals with the assessment of respiratory disorder using speech parameters by comparing different classifiers on the same dataset. As per the literature survey, the main drawback was the use of various datasets which makes the comparison even more complicated.

Therefore, for the comparison of classifiers, there is a need of using the same dataset.

In this article, we have compared 5 classifiers on the basis of multifold cross-validation and the hold-out method, and 15 features were extracted from the speech of each of the 20 participants.

Summary of the study contribution:

– in this study, we have proposed a speech signal-based detection of affected speech. Different speech features were extracted from the speech signal and evaluated using classification techniques to detect abnormalities in the speech parameters;
– we also implemented and evaluated different machine learning classifiers capable of differentiating healthy speech and affected speech.

2. Materials and methods

This section presents different steps involved in the systematic classification of speech features.

2.1. Data collection

The dataset comprises speech samples of 20 individuals aged between 24–65 years, 10 healthy (6 males and 4 females), and 10 patients (10 males and 0 females). All the participants have given their written consent.

Samples were recorded using Goldwave software with the sampling frequency of 11 025 Hz by a microphone located 2–3 cm in front of the participant's lips. The participants were asked to repeat specified words in Hindi while sitting and to adopt pitch and loudness with which they were usually comfortable. Each individual recorded the speech for two minutes in a continuous manner. Only one recording was obtained from each patient.

The database consisting of a sustained phonation was created. The input signal waveforms of healthy and affected people' speech are shown in Fig. 1.

a)

b)

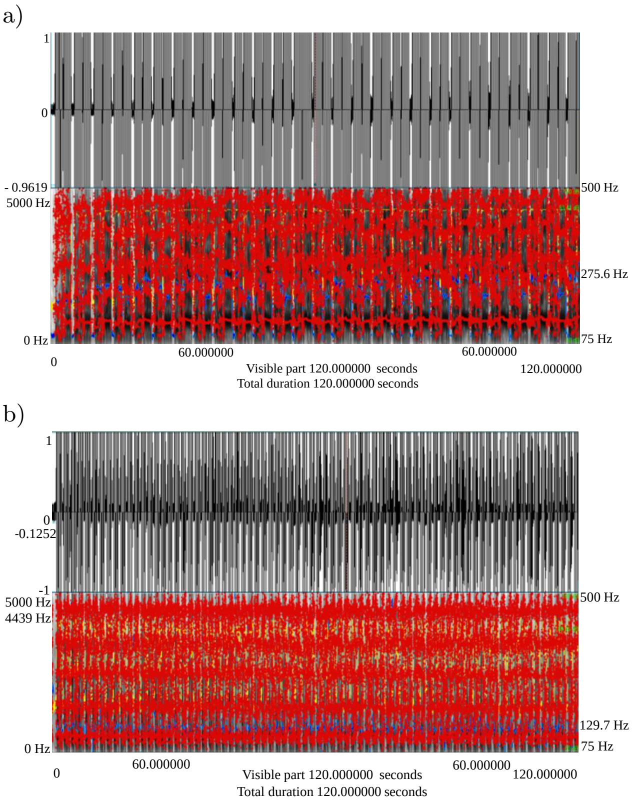

Fig. 1. (a) Healthy person's input signal waveform; (b) affected person's input signal waveform.

From Fig. 1 it was observed that for the healthy person, the waveform was uniform but for the affected person, the waveform contains deformities.

For the lung assessment, a spirometry test had been performed before speech recording for all participants.

This test estimates the amount of air that can be breathed in and breathed out of the lungs, as well as if the air can be blown out of the lungs, fast and easily. In spirometry, participants were asked to inhale deeply and hold for 6 seconds then exhale completely. Thus, the forced expiratory volume in 1 second (FEV1), defined as the amount of air that can be forced from the lungs in one second, and peak expiratory flow rate (PEFR) which measures how much air flows through the bronchi and displays the level of obstruction in the lungs, were recorded.

All the participants were subjected to a spirometry test before categorising them as healthy or affected under clinical supervision. The values of FEV1 and PEFR were used as the gold standard to differentiate the two groups. A participant was considered affected if the values of FEV1 and PEFR were observed to be less than 60.

The vital information such as age, gender, height, weight, occupation, and medical history was also noted down and included in the final analysis other than spirometry.

2.2. Feature extraction

A set of 15 speech features, namely, formant frequencies, F1, F2, and F3, pitch, intensity, jitter (rap, %), mean autocorrelation, jitter (local, %), mean NHR, shimmer (local, %), mean HNR, amplitude mean (Pascal), total energy ($\text{Pa}^2 \cdot \text{s}$), mean power (intensity), and standard deviation in a channel (Pascal) was extracted with the help of PRAAT software.

The following speech parameters are explained as follows:

– Formant frequencies: speech producing different frequency components of the sound signal is called formant frequencies.
– Pitch: pitch is defined as the ordering of sound property on a scale that is frequency-related. Thus, the relative highness or lowness of a tone is considered pitch.
– Intensity: the power of sound per unit area is known as intensity. Also defined as the amplitude of the vibrations that affect loudness.
– Jitter: jitter may be defined as frequency parameters variation from cycle to cycle.
Relative jitter or local jitter is defined as the ratio between the average differences between consecutive periods, relative to the overall average period. It is given in percentage:

$$\text{Jitter (relative)} = \frac{\frac{1}{N-1}\sum\limits_{i=1}^{N-1}|T_i - T_{i-1}|}{\frac{1}{N}\sum\limits_{i=1}^{N}T_i} \cdot 100, \quad (1)$$

where T_i is extracted F0 period lengths, N is number of extracted F0 periods.

Jitter (rap): the average absolute difference between a period and the average of it and its two neighbors, divided by the average period is defined as jitter (rap). It is expressed as a percentage:

$$\text{Jitter (rap)} = \frac{\frac{1}{N-1}\sum\limits_{i=1}^{N-1}\left|T_i - \left(\frac{1}{3}\sum\limits_{n=i-1}^{i+1}T_n\right)\right|}{\frac{1}{N}\sum\limits_{i=1}^{N}T_i} \cdot 100. \quad (2)$$

– Shimmer: it may be represented as the parameters associated with the variation of the amplitude of the sound wave.
Shimmer relative: the ratio between the average absolute difference between the amplitudes of consecutive periods and the average amplitude is defined as shimmer relative, given in percentage:

$$\text{Shimer (relative)} = \frac{\frac{1}{N-1}\sum\limits_{i=1}^{N-1}|A_i - A_{i+1}|}{\frac{1}{N}\sum\limits_{i=1}^{N}A_i} \cdot 100, \quad (3)$$

where A_i is extracted peak-to-peak amplitude data, N is number of extracted fundamental frequency periods.
– Mean autocorrelation: the relationship between the current values of variables and their past values is measured by autocorrelation. The correlation coefficient is usually denoted ρ. For variables, x and y, each contains N values:

$$\rho = \frac{\sum\limits_i (x_i - \mu_x)(y_i - \mu_y)}{N\sigma_x\sigma_y}, \quad (4)$$

where the means of x and y are given by μ_x and μ_y, and their standard deviations are given as σ_x and σ_y.
– Harmonic to noise ratio (HNR): the periodic components of speech sound divided by non-periodic components is represent by harmonic to noise ratio:

$$\text{HNR} = 10 \cdot \log_{10}\frac{AC_V(T)}{AC_V(0) - AC_V(T)}, \quad (5)$$

where the autocorrelation coefficient consisting of all signal energy at the origin is given by $AC_V(0)$, and the autocorrelation component related to the fundamental period is given by $AC_V(T)$.
– Noise to harmonic ratio (NHR): hoarseness can be measured effectively by noise to harmonic ratio:

$$\text{NHR} = 1 - \text{autocorrelation}. \quad (6)$$

– Amplitude mean: it is the amplitude of the vibrations that affects loudness which is the size of oscillations of the vocal folds.

2.3. Statistical significant analysis

Now, to determine the statistical significance of extracted features, statistically significant analysis using Statistical Package for Social Sciences (SPSS) was applied. Each study has a confidence level of 95% and a p-value of <0.05 considering being statistically significant.

2.4. Classification

Classification is a process in which the class of given data points is predicted and along with targets, input data are also provided. This type of classification comes under the category of supervised learning.

Thus, to understand the given input variables related to the class, a classifier utilises some training data.

There are several classification algorithms varying in the nature and application of available data.

In the present study, the following classification algorithms were evaluated:

- KNN with all the kernels,
- SVM with all the kernels,
- decision tree,
- logistic regression,
- linear discriminant.

As these algorithms represent a variety of classifiers' algorithms, they were chosen (Kuncheva, 2014) and also some of them had performed well in previous studies (Caruana, Niculescu-Mizil, 2006).

After all the statistically significant features were obtained, they were used for the classification of healthy and affected speech. For this, different supervised machine learning techniques were applied, explained as follows:

- Decision tree: as the name specifies, the classification or regression models are built in the form of a tree structure in the decision tree method. The method applied an if-then rule set that is both mutually exclusive and exhaustive for the classification. One at a time the training data are used for learning the rule sequentially.
- SVM: for classification and regression problems, support vector machines are widely used. Figure 2 represents the SVM in a two-dimensional space. By constructing a hyperplane, it separates two classes of a sample to distinguish class members from non-members (Byun, Lee, 2002). A hyperplane is constructed as the decision plane in SVM, separating the positive (+1) and negative (−1) classes with the largest margin. The maximum margin of separation between the two classes is an optimal hyperplane, where the margin is the sum of the distances from the hyperplane to the closest data points of each of the two classes. These closest data points are called support vectors (SVs). The optimal separating hyperplane is represented by the solid line in Fig. 2.

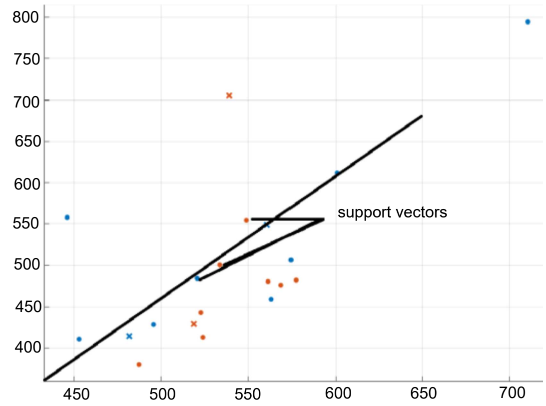

Fig. 2. Representation of SVM in a two-dimensional space.

In a variety of classification and regression theories, SVM has been successfully used (Sapankevych, Sankar, 2009):

- Logistic regression: to model the conditional probability, the logistic function is used by the statistical model and this is known as logistic regression. Basically when the target variable's value is categorical then this classification algorithm is used. Thus, most commonly used when the data in question has binary output.
  The formula is given:
  $$P = \frac{1}{1 + e^{-(a+bX)}}, \tag{7}$$
  where P is the probability of 1 (the proportion of 1 s), e is the base of the natural algorithm, a and b are parameters of the models, and X is the independent variable related to the logistic curve.
- Linear discriminant: linear discriminant analysis is a simple and effective supervised classification method, used to create machine learning models (Amaral *et al.*, 2012). It is used for modelling differences in groups, i.e., separating two or more classes.
- KNN: KNN method that uses data and classifies new data points based on similarity measures is considered to be the simplest method applied for regression and classification problems (Amaral *et al.*, 2013). Classification is done by a maximum vote from its neighbors. KNN was briefly defined in the previous works, as KNN calculates the distance between data points. For this, the simple Euclidean distance formula is generally used:
  $$\mathrm{d}(p,q) = \mathrm{d}(q,p) = \sqrt{\sum_{i=1}^{n} (\mathbf{q}_i - \mathbf{p}_i)^2}, \tag{8}$$

where p and q are two points in Euclidean n-space, $\mathbf{q}_i$ and $\mathbf{p}_i$ are Euclidean vectors, starting from the origin of the space (initial point), n is n-space.

The block diagram and overall proposed strategy are depicted in Fig. 3.

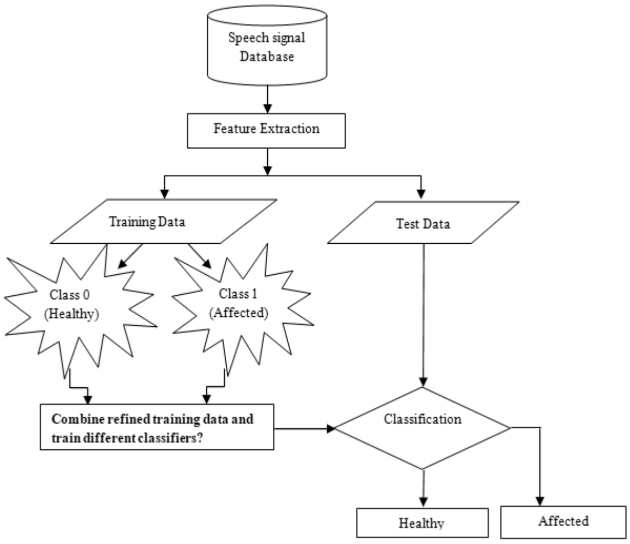

Fig. 3. Flowchart of the proposed classification approach.

The dataset has been divided into 67% training and 33% testing parts in the case of the holdout method, while in the case of the multifold cross-validation process both 5-fold and 10-fold have been used. After classification, the healthy speech was distinguished from the affected speech signals.

2.5. *Performance evaluation*

2.5.1. *Data division protocol*

Two groups of the dataset were formed to evaluate the performance of the proposed classifier models. One group was used for training purposes while the other was used for testing, and k-fold cross-validation and the holdout were applied as two data division protocols. The holdout is the most common method out of the several existing in which the given dataset has two groups divided randomly. The train set will be used to train the data set, and the unseen test data will be used to test its predictive power. 67% of the samples were used for training while 33% were used for testing. The common problem that generally occurs in most of the models in machine learning is over-fitting. So, to verify that the model is not overfit, k-fold cross-validation can be conducted in which random partition of the data set into k manually exclusive groups each approximately of the same size is made. For testing, one is kept, while for training others are used (Refaeilzadeh *et al.*, 2009). The experiments were conducted with the 5-fold and 10-fold cross-validation.

2.5.2. *Performance measures*

After training the data, the testing was performed and the following performance parameters were used for evaluation. Accuracy, recall, specificity, true positive rate, sensitivity, false positive rate, precision, and the area under the receiver operating characteristic (ROC), and the area under curve (AUC) are some well known performance criteria (Fawcett, 2006). In this paper, we have selected accuracy, sensitivity, specificity, and AUC for ROC curves as they are generally applied in medical diagnoses:

– Accuracy: accuracy is obtained by dividing all the correct predictions by the total number of predictions:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{9}$$

where TP – true positive, TN – true negative, FP – false positive, FN – false negative.

– Sensitivity: sensitivity is calculated by dividing true positive by true positive plus false negative:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{10}$$

– Specificity: true negative divided by the sum of true negative and false positive is defined as specificity:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \tag{11}$$

– Area under curve (AUC): to evaluate the performance, a single value metric is used, known as area under ROC curve, plotted between true positive rate (TPR) on the y-axis and false positive rate (FPR) on the x-axis at the various thresholds.

3. Results and discussion

Table 1 presents an independent sample t-test applied to the database of 15 features. The detailed statistics of various features were divided into two classes, i.e., class 0 (healthy) and class 1 (affected) using SPSS software.

It has been observed that t-values and df-values did not provide any significant variances but sig (2-tailed) provided values of mean autocorrelation to be 0.011 and 0.012 when equal variance was assumed and when it was not assumed, respectively, as presented in Table 1. Similarly, mean NHR values were achieved to be 0.009 and 0.010, and mean HNR values were 0.000 for both the cases, i.e., when equal variance was assumed and when it was not assumed. This test shows that only the mean autocorrelation, mean NHR, and mean HNR were statistically evident in the healthy and affected speech (p-value < 0.05 and confidence interval of 95%), while the rest features were not statistically

Table 1. t-test for equality means.

Independent samples test		
Speech features	Conditions	t-test for equality of means sig (2-tailed)
F1	Equal variances assumed	0.838
	Equal variances not assumed	0.839
F2	Equal variances assumed	0.137
	Equal variances not assumed	0.145
F3	Equal variances assumed	0.425
	Equal variances not assumed	0.426
Pitch	Equal variances assumed	0.448
	Equal variances not assumed	0.451
Intensity	Equal variances assumed	0.973
	Equal variances not assumed	0.973
Jitter (local)	Equal variances assumed	0.190
	Equal variances not assumed	0.190
Jitter (rap)	Equal variances assumed	0.319
	Equal variances not assumed	0.333
Shimmer (local)	Equal variances assumed	0.365
	Equal variances not assumed	0.367
mean Auto Correlation	Equal variances assumed	0.011
	Equal variances not assumed	0.012
mean NHR	Equal variances assumed	0.009
	Equal variances not assumed	0.010
mean HNR	Equal variances assumed	0.000
	Equal variances not assumed	0.000
Amplitude mean [Pa]	Equal variances assumed	0.617
	Equal variances not assumed	0.621
Total energy [$Pa^2 \cdot s$]	Equal variances assumed	0.672
	Equal variances not assumed	0.672
mean Power (intensity) in air [dB]	Equal variances assumed	0.971
	Equal variances not assumed	0.971
Standard deviation in the channel [Pa]	Equal variances assumed	0.763
	Equal variances not assumed	0.764

significant. Hence these three features were considered for further classification.

Four performance measures, namely: accuracy, sensitivity, specificity, and AUC were used for evaluation under three data division protocols, namely: 5-fold, 10-fold, and the holdout. MATLAB software was used to perform this step.

Table 2 shows the performance of various classifier models using 5-fold, 10-fold, and the holdout cross-validation when 15 features were considered. It was observed that in 5-fold and 10-fold, the decision tree achieved the highest values of classification accuracy of 90%. The other performance measures, namely, sensitivity and specificity were found to be 90%, and AUC was 0.82. Whereas logistic regression and linear discriminant achieved the lowest values. Cubic KNN also shows 90% classification accuracy in the case of the 5-fold method. Also, coarse KNN achieved the least value of classification accuracy and AUC, i.e., 0.5, but achieved the highest value for specificity, i.e., 100%, and zero value for sensitivity.

It was found that in the holdout method, decision tree, linear discriminant, all the kernels of SVM (except quadratic SVM), and weighted KNN achieved 100% sensitivity. Zero specificity was observed in fine Gaussian SVM and fine KNN. Also, zero sensitivity was again observed in the case of coarse KNN along with 50% of accuracy and 86% of specificity. The decision tree achieved the highest accuracy of 83%.

Table 3 shows the performance evaluation of three significant features obtained by statistical analysis using SPSS, namely: mean NHR, mean HNR, and mean

Table 2. Performance of various classifiers using 15 features under different data division protocols.

Data division protocol	Classification techniques	Performance measures			
		Accuracy [%]	Sensitivity [%]	Specificity [%]	AUC
5-fold	Decision tree	90	90	90	0.82
	Linear discriminant	45	40	50	0.45
	Logistic regression	45	40	50	0.45
	Linear SVM	75	80	70	0.70
	Quadratic SVM	65	70	60	0.67
	Cubic SVM	45	40	50	0.68
	Fine Gaussian SVM	45	60	30	0.41
	Medium Gaussian SVM	80	80	80	0.80
	Coarse Gaussian SVM	75	80	70	0.75
	Fine KNN	60	60	60	0.60
	Medium KNN	85	90	80	0.89
	Coarse KNN	50	0	100	0.50
	Cosine KNN	85	80	90	0.86
	Cubic KNN	90	90	90	0.89
	Weighted KNN	75	90	60	0.83
10-fold	Decision tree	90	90	90	0.82
	Linear discriminant	55	50	60	0.49
	Logistic regression	40	50	30	0.42
	Linear SVM	65	70	60	0.66
	Quadratic SVM	75	90	60	0.82
	Cubic SVM	60	60	60	0.67
	Fine Gaussian SVM	50	80	20	0.41
	Medium Gaussian SVM	80	80	80	0.81
	Coarse Gaussian SVM	80	80	80	0.49
	Fine KNN	55	60	50	0.55
	Medium KNN	70	90	50	0.84
	Coarse KNN	50	0	100	0.50
	Cosine KNN	80	80	80	0.84
	Cubic KNN	70	80	60	0.78
	Weighted KNN	75	90	60	0.81
Holdout	Decision tree	83	100	71	0.83
	Linear discriminant	67	67	75	0.67
	Logistic regression	67	67	75	0.67
	Linear SVM	67	100	60	0.56
	Quadratic SVM	50	67	67	0.56
	Cubic SVM	67	100	60	0.78
	Fine Gaussian SVM	50	100	0	0.67
	Medium Gaussian SVM	67	100	60	0.89
	Coarse Gaussian SVM	67	100	60	0.44
	Fine KNN	50	100	0	0.50
	Medium KNN	67	67	75	0.67
	Coarse KNN	50	0	86	0.50
	Cosine KNN	50	67	67	0.56
	Cubic KNN	50	33	80	0.61
	Weighted KNN	67	100	60	0.56

Table 3. Performance of various classifiers using 3 significant features under different data division protocols.

Data division protocol	Classification techniques	Performance measures			
		Accuracy [%]	Sensitivity [%]	Specificity [%]	AUC
5-fold	Decision tree	90	90	90	0.82
	Linear discriminant	75	70	80	0.76
	Logistic regression	75	70	80	0.69
	Linear SVM	85	90	80	0.87
	Quadratic SVM	80	80	80	0.84
	Cubic SVM	70	70	70	0.67
	Fine Gaussian SVM	80	70	90	0.89
	Medium Gaussian SVM	80	80	80	0.85
	Coarse Gaussian SVM	85	90	80	0.85
	Fine KNN	85	90	80	0.85
	Medium KNN	85	100	80	0.79
	Coarse KNN	50	0	100	0.50
	Cosine KNN	85	90	80	0.84
	Cubic KNN	85	90	80	0.79
	Weighted KNN	85	90	80	0.85
10-fold	Decision tree	90	90	90	0.82
	Linear discriminant	80	80	80	0.92
	Logistic regression	75	70	80	0.81
	Linear SVM	85	90	80	0.86
	Quadratic SVM	85	90	80	0.81
	Cubic SVM	70	70	70	0.63
	Fine Gaussian SVM	75	70	80	0.87
	Medium Gaussian SVM	80	80	80	0.84
	Coarse Gaussian SVM	85	90	80	0.85
	Fine KNN	80	80	80	0.80
	Medium KNN	85	90	80	0.85
	Coarse KNN	50	0	100	0.50
	Cosine KNN	85	90	80	0.90
	Cubic KNN	85	90	80	0.82
	Weighted KNN	85	90	80	0.84
Holdout	Decision tree	100	100	100	1.00
	Linear discriminant	80	67	100	1.00
	Logistic regression	80	67	100	0.83
	Linear SVM	100	100	100	1.00
	Quadratic SVM	100	100	100	1.00
	Cubic SVM	80	100	50	0.50
	Fine Gaussian SVM	80	67	100	0.83
	Medium Gaussian SVM	100	100	100	1.00
	Coarse Gaussian SVM	80	100	67	1.00
	Fine KNN	80	67	100	0.83
	Medium KNN	60	33	100	1.00
	Coarse KNN	40	0	100	0.50
	Cosine KNN	100	100	100	1.00
	Cubic KNN	60	33	100	1.00
	Weighted KNN	80	67	100	1.00

autocorrelation that showed significant variance. The same procedure was adopted and the result was shown.

Table 3 summarises the performance measures of various classifier methods when only three significant features were considered using different data division protocols. It was found that 90% of classification accuracy was shown by a decision tree in 5-fold as well as 10-fold cross-validation. The other performance measures, namely, sensitivity and specificity were 90%, and AUC was 0.82. Again, coarse KNN showed the smallest value of classification accuracy, i.e., 50%, and AUC to be 0.5 but gave the highest value for specificity, i.e., 100%, and zero value for sensitivity.

Whereas in the holdout method it was found that decision tree, linear SVM, quadratic SVM, medium Gaussian SVM, and cosine KNN have shown classification accuracy, sensitivity, and specificity to be 100%, and AUC to be 1. Specificity was found to be 100% in almost all the classifiers except cubic SVM and coarse Gaussian SVM. Coarse KNN achieved zero sensitivity, 40% accuracy, 0.5 AUC along with 100% specificity.

On analysing Tables 2 and 3, it was observed that sensitivity, accuracy, specificity, and AUC were found to be high in the case of the decision tree as compared to the rest of the classifiers. Similarly, the holdout method was found to be the best performer out of all the three data division protocols considering both 15 features and three significant features.

It was also observed that, out of all the kernels of KNNs, coarse KNN, and out of all the kernels of SVMs, fine Gaussian SVM was giving poor performance throughout the experiment.

The diagnostic capability of classifiers and features can be determined using TPR and TNR. A plot between TPR and TNR is called receiver operating characteristics (ROC). Figure 4 shows the result of the

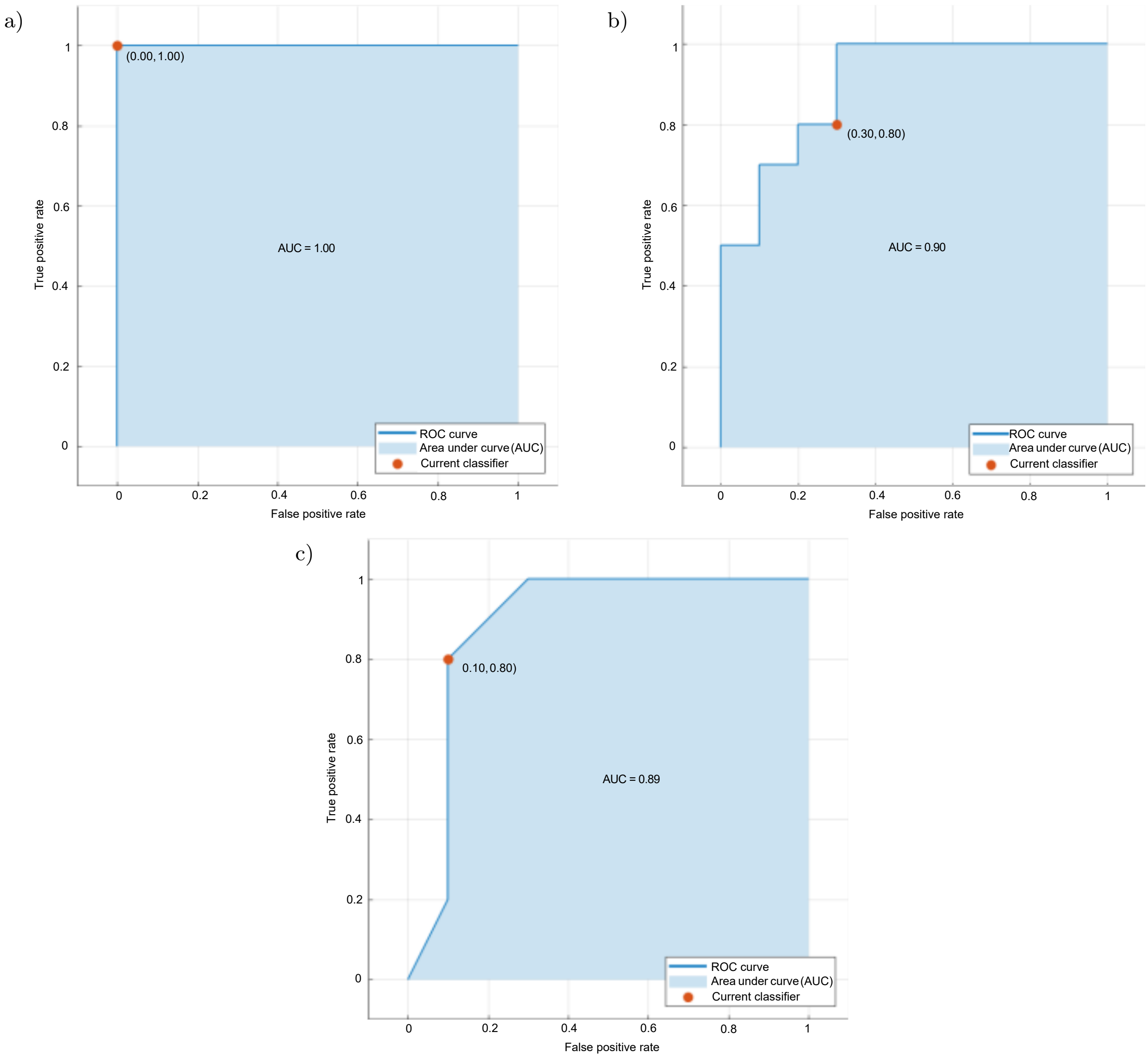

Fig. 4. ROC plots for various classifiers: a) decision tree (hold and method); b) cosine KNN (10-fold); c) cubic KNN (5-fold).

best three ROC analyses. It was observed that the AUC of 1 was obtained by the holdout method for the decision tree and the AUC of 0.90 and 0.89 were achieved by k-fold methods for KNN classifiers and verified in the results of Tables 2 and 3.

The best three ROC curves observed during the analysis were represented in Fig. 4.

4. Conclusion

It has been observed that on comparing different classifiers for all features, accuracy, specificity, sensitivity, and area under the curve performance measures for 5-fold, 10-fold cross-validation, and the holdout method, the decision tree achieved 90–100% classification accuracy. Further, SVMs and KNNs achieved the lowest accuracy between 40 and 70%. The holdout method had given a promising result.

Similarly, on comparing the accuracy with three significant features almost the same result was shown by both the 5-fold and 10-fold cross-validation methods for all the various classifiers. For the decision tree, it was again 90%, for SVM and KNN it was found to be 90–100%. But for logistic regression and linear discriminant, the performance improved to 80–90%. Again, the holdout method had given almost perfect results.

Thus, it is concluded that using different machine learning techniques, a comparative analysis of classifiers shows that the decision tree was effective as classification accuracy achieved 90% along with the holdout data division protocol for classification of speech of healthy and affected individuals.

Acknowledgments

The authors express their sincere thanks to the Director and Management of the research center, Shri Shankara College of Engineering and Technology, Bhilai for permitting to use the facilities of their Institute.

The authors extend their appreciation to the Dean and the supporting staff of the Pulmonary department of the O.P Jindal Fortis Hospital and Research Center for permitting to use their facilities for creating the database.

References

1. Alghowinem S. *et al.* (2013), A comparative study of different classifiers for detecting depression from spontaneous speech, [in:] *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8022–8026, doi: 10.1109/ICASSP.2013.6639227.
2. Amaral J.L., Lopes A.J., Jansen J.M., Faria A.C., Melo P.L. (2012), Machine learning algorithms and forced oscillation measurements applied to the automatic identification of chronic obstructive pulmonary disease, *Computer Methods and Programs in Biomedicine*, **105**(3): 183–193, doi: 10.1016/j.cmpb.2011.09.009.
3. Amaral J.L., Lopes A.J., Jansen J.M., Faria A.C., Melo P.L. (2013), An improved method of early diagnosis of smoking-induced respiratory changes using machine learning algorithms, *Computer Methods and Programs in Biomedicine*, **112**(3): 441–454, doi: 10.1016/j.cmpb.2013.08.004.
4. Byun H., Lee S.W. (2002), Applications of support vector machines for pattern recognition: A survey, [in:] *International Workshop on Support Vector Machines*, pp. 213–236, Springer, Berlin, Heidelberg, doi: 10.1007/3-540-45665-1_17.
5. Calverley P.M.A. (2020), Defining airflow obstruction: More data, further clarity, *American Journal of Respiratory and Critical Care Medicine*, **202**(5): 649–650, doi: 10.1164/rccm.202005-1551ED.
6. Caruana R., Niculescu-Mizil A. (2006), An empirical comparison of supervised learning algorithms, [in:] *Proceedings of the 23rd International Conference on Machine Learning*, pp. 161–168, doi: 10.1145/1143844.1143865.
7. Chun K.S. *et al.* (2020), Towards passive assessment of pulmonary function from natural speech recorded using a mobile phone, [in:] *2020 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 1–10, doi: 10.1109/PerCom45495.2020.9127380.
8. Dixit, Mittal V., Sharma Y. (2014), Voice parameter analysis for the disease detection, *IOSR Journal of Electronics and Communication Engineering*, **9**(3): 48–55.
9. Dogan M., Eryuksel E., Kocak I., Celikel T., Sehitoglu M.A. (2007), Subjective and objective evaluation of voice quality in patients with asthma, *Journal of Voice*, **21**(2): 224–230, doi: 10.1016/j.jvoice.2005.11.003.
10. Fawcett T. (2006), An introduction to ROC analysis, *Pattern Recognition Letters*, **27**(8): 861–874, doi: 10.1016/j.patrec.2005.10.010.
11. Gore S.M., Salunke M.M., Patil S.A., Kemalkar A.K. (2020), Disease detection using voice analysis, *International Research Journal of Engineering and Technology*, **7**(5): 7655–7659.
12. Gürbüz E., Kiliç E. (2014), A new adaptive support vector machine for diagnosis of diseases, *Expert Systems*, **31**(5), 389–397, doi: 10.1111/exsy.12051.
13. Halpin D.M. *et al.* (2021), Global initiative for the diagnosis, management, and prevention of chronic obstructive lung disease. The 2020 GOLD science committee report on COVID-19 and chronic obstructive pulmonary disease, *American Journal of Respiratory and Critical Care Medicine*, **203**(1): 24–36, doi: 10.1164/rccm.202009-3533SO.
14. Jain D., Singh V. (2018), Feature selection and classification systems for chronic disease prediction: A re-

view, *Egyptian Informatics Journal*, **19**(3): 179–189, doi: 10.1016/j.eij.2018.03.002.

15. Kocsis O. *et al.* (2017), Assessing machine learning algorithms for self-management of asthma, [in:] *2017 E-Health and Bioengineering Conference (EHB)*, pp. 571–574, doi: 10.1109/EHB.2017.7995488.
16. Kuncheva L.I. (2014), *Combining Pattern Classifiers: Methods and Algorithms*, John Wiley & Sons.
17. Mohamed E.E., El Maghraby R.A. (2014), Voice changes in patients with chronic obstructive pulmonary disease, *Egyptian Journal of Chest Diseases and Tuberculosis*, **63**(3), 561–567, doi: 10.1016/j.ejcdt.2014.03.006.
18. Refaeilzadeh P., Tang L., Liu H. (2009), Cross-validation, [in:] *Encyclopedia of Database Systems*, Liu L., Özsu M. [Eds.], Springer, doi: 10.1007/978-0-387-39940-9_565.
19. Saloni, Sharma R.K., Gupta A.K. (2014), Disease detection using voice analysis: A review, *International Journal of Medical Engineering and Informatics*, **6**(3): 189–209, doi: 10.1504/IJMEI.2014.063173.
20. Sapankevych N.I., Sankar R. (2009), Time series prediction using support vector machines: A survey, *IEEE Computational Intelligence Magazine*, **4**(2): 24–38, doi: 10.1109/MCI.2009.932254.
21. Sonu, Sharma R.K. (2011), Disease detection using analysis of voice parameters, [in:] *5th IEEE International Conference on Advanced Computing and Communication Technologies (ICACCT-2011)*, Choudhary R.K., Verma M., Saini S. [Eds.], pp. 416–420.
22. Teixeira J.P., Fernandes P.O. (2014), Jitter, shimmer and HNR classification within gender, tones and vowels in healthy voices, *Procedia Technology*, **16**: 1228–1237, doi: 10.1016/j.protcy.2014.10.138.
23. Tsang K.C., Pinnock H., Wilson A.M., Shah S.A. (2020), Application of machine learning to support self-management of asthma with mHealth, [in:] *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 5673–5677, doi: 10.1109/EMBC44109.2020.9175679.
24. Walia G.S., Sharma R.K. (2016), Level of asthma: mathematical formulation based on acoustic parameters, [in:] *2016 Conference on Advances in Signal Processing (CASP)*, pp. 24–27, doi: 10.1109/CASP.2016.7746131.
25. Wiechern B., Liberty K.A., Pattemore P., Lin E. (2018), Effects of asthma on breathing during reading aloud, *Speech, Language and Hearing*, **21**(1): 30–40, doi: 10.1080/2050571X.2017.1322740.