

International Journal of Intelligent Systems Technologies and Applications

ISSN online: 1740-8873 - ISSN print: 1740-8865
<https://www.inderscience.com/ijista>

Spectral descriptors for the assessment of vocal fold nodules and feature optimisation using MRMR algorithm

Jennifer C. Saldanha, Malini Suvarna, Dayakshini Sathish, Cynthia Santhmayor, Rohan Pinto

DOI: [10.1504/IJISTA.2025.10072833](https://doi.org/10.1504/IJISTA.2025.10072833)

Article History:

Received:	24 August 2022
Last revised:	11 August 2024
Accepted:	10 September 2024
Published online:	30 September 2025

Spectral descriptors for the assessment of vocal fold nodules and feature optimisation using MRMR algorithm

Jennifer C. Saldanha*

Department of Electronics and Communication,
St. Joseph Engineering College,
Vamanjoor, Mangaluru, India
and
Affiliated to: VTU Belagavi, India
Email: jennifers@sjec.ac.in
*Corresponding author

Malini Suvarna

Department of Electronics and Communication,
Tontadarya College of Engineering,
Gadag, India
and
Affiliated to: VTU Belagavi, India
Email: suvarnamalini50@gmail.com

Dayakshini Sathish

Department of Electronics and Communication,
St. Joseph Engineering College,
Vamanjoor, Mangaluru, India
and
Affiliated to: VTU Belagavi, India
Email: dayakshini@sjec.ac.in

Cynthia Santhmayor

Department of Speech,
Father Muller College of Speech and Hearing,
Kankanady, Mangalore, India
Email: cynthihc@gmail.com

Rohan Pinto

Department of Electronics and Communication,
St. Joseph Engineering Collge,
Mangalore, India
and

Affiliated to: VTU Belagavi, India
Email: rohanp@sjec.ac.in

Abstract: Objective assessment of voice in vocal nodules from the spectral descriptive features is discussed in this study. Further, MATLAB feature engineering is explored to automate the process of feature engineering. The performance of a set of optimisable classifiers such as decision tree, bagged trees, Naive Bayes, linear and quadratic support vector machines was evaluated on feature engineered dataset. The decision tree outperformed all other classifiers with an accuracy of 84.2% for engineered features. Spectral centroid obtained highest ranking using maximum relevance and minimum redundancy feature ranking method and found to be most appropriate for classification. Harmonic ratio, harmonic to noise ratio, shimmer variants, and spectral centroid features obtained a significant amount of correlation with the perceived degree of hoarseness. Among these features, spectral centroid is found to be strongly negatively correlated, hence can be effectively used as a quantitative indicator to measure the level of severity of the pathologic voice.

Keywords: maximum relevance minimum redundancy; MRMR; harmonic to noise ratio; HNR; noise to harmonic ratio; NHR; cepstral peak prominence; CPP; correlation feature selection; CFS; support vector machines; SVMs; artificial neural network; ANN.

Reference to this paper should be made as follows: Saldanha, J.C., Suvarna, M., Sathish, D., Santhmayor, C. and Pinto, R. (2025) ‘Spectral descriptors for the assessment of vocal fold nodules and feature optimisation using MRMR algorithm’, *Int. J. Intelligent Systems Technologies and Applications*, Vol. 23, No. 3, pp.337–362.

Biographical notes: Jennifer C. Saldanha is an academician with 14 years of teaching experience. She is currently working as an Assistant Professor in the Department of Electronics and Communication, St. Joseph Engineering College, Mangaluru. Her MTech specialisation is Digital Electronics and Communication, she is an active researcher in the field of speech processing and machine learning.

Malini Suvarna received her PhD in the field of Image Processing at Manipal Academy of Higher Education, Manipal. She is an academician with 23 years teaching experience. Currently she is working as a Professor in the Department of Electronics and Communication, Tontadarya College of Engineering, Gadag. She is an active researcher in the field of digital image processing. She has published research papers in national and international journals, conference proceedings as well as chapters of books. She has organised several workshops and faculty development programs.

Dayakshini Sathish obtained her PhD from Manipal Academy of Higher Education, Manipal in the field of Biomedical Image Processing. She is an academician with 22 years of teaching and research experience. She is currently working as a Professor and the Head of the Department of Electronics and Communication Engineering, St. Joseph Engineering College, Mangaluru. She is an active researcher in the field of medical image/signal processing and published research articles in international journals indexed by Web of Science and Scopus. Also, she is a reviewer of reputed journals published by IEEE, Elsevier, Springer etc. She has organised several workshops, faculty development programs and served as resource person.

Cynthia Santhmayor is an academician with nine years of teaching experience with specialisation in audiology and speech language pathology. Currently, she is working as a Professor and the Head in the Department of Speech Language Pathology, Father Muller College of Speech and Hearing, Kankanady, Mangalore. She is an active researcher in the field of speech language pathology. She has published research papers in national and international journals, conference proceedings as well as chapters of books.

Rohan Pinto obtained her PhD from Manipal Academy of Higher Education, Manipal in the field of VLSI Signal Processing. He is an academician with 14 years of teaching and research experience. He is currently working as a Professor in the Department of Electronics and Communication Engineering, St. Joseph Engineering College, Mangaluru. He is an active researcher in the field of signal processing and published research articles in international journals indexed by Scopus. Also, he is a reviewer of reputed journals published by IEEE and Springer.

This paper is a revised and expanded version of a paper entitled [title] presented at [name, location and date of conference].

1 Introduction

Voice pathology can be caused by infection of the tissue, vocal abuse, neurological variations, and organic changes at the larynx (Harar et al., 2018). Any change in the mass, size, tension, and elasticity of the vocal folds due to pathology results in irregular vibration patterns and introduces noise in the acoustic signal. The voice produced by disordered vocal fold vibration exhibits hoarse, harsh, and breathy voice quality (Martin et al., 1995; Lopes et al., 2014). Voice pathology assessment is carried out by both subjective and objective assessment. Subjective assessment includes perceptual and visual evaluation of vocal folds. The visual inspection includes laryngoscopy, endoscopy, and stroboscopy. Whereas the perceptual evaluation includes listening to the patient's voice and rating the severity based on the clinical rating scale. Subjective methods are costly, time-consuming, and dependent on the medical practitioner who carries out the test. The objective evaluation is a non-invasive technique where a computer-aided diagnosis is carried out. Hence, it is more accurate and free from subjective bias.

Objective assessment of voice pathology has been intensively studied by the researchers on the databases such as Massachusetts Eye and Ear Infirmary (MEEI)

Database, Saarbrücken Voice Database (SVD), and Arabic Voice Pathology Database (AVPD) (Al-Nasheri et al., 2017a). Most of the researchers used conventional feature extraction methods such as jitter, shimmer, Mel-frequency cepstral coefficients, and linear prediction cepstral coefficients for the analysis and classifiers such as k-nearest neighbour (k-NN), Gaussian mixture model (GMM), and artificial neural network (ANN) for classifying the pathological sample from the healthy ones. The classification results may vary due to different database, features extracted, and classifiers that are used in those papers. The proposed work is focused on the analysis and detection of vocal fold nodules which belong to the structural organic category of voice pathology which occurs due to physical change in the laryngeal anatomy. A recent study has indicated the prevalence of voice disorder in the Dakshina Kannada region to be 21.4%. In addition to this, the study has also revealed the presence of vocal fold nodules to a larger extent (Dodderi et al., 2018).

Vocal fold nodules are small lesions developed at the midpoint of the vocal folds (Naunheim and Carroll, 2017). The main cause for vocal fold nodules is overuse or misuse of vocal apparatus. This voice pathology is mostly observed in children, adult singers, and teachers. School children and women in the age group of 20 to 50 years are more likely prone to vocal fold nodules. The other causes for vocal fold nodules include smoking, use of alcohol, sinusitis, and allergies. The typical voice-related issues associated with vocal fold nodules are frequent voice breaks, decreased range of pitch, and vocal fatigue. The subjects with the problem of vocal nodules are usually treated with voice therapy and behavioural changes. Surgery is reserved for the cases where there is no improvement in the voice quality from voice therapy. The mean values of acoustic features such as jitter (%), harmonic to noise ratio (HNR), and voice handicap index are used for measuring the level of dysphonia in case of voice affected by the nodules (Schindler et al., 2012).

The current work differs from the already published work in the literature as it uses a private database from the local hospital. Spectral features which are used widely for audio signal processing applications are used here for pathological speech analysis. The current study also explores the role of feature engineering in increasing the accuracy of classification. The classification results of the optimisable decision tree, bagged trees, Naive Bayes, neural network, linear and quadratic support vector machine (SVM) (Bajpai et al., 2023) classifiers are compared to find the classifier which gives maximum accuracy for the test dataset. The maximum relevance minimum redundancy (MRMR) algorithm is used to select the most relevant feature for characterising the dysphonic voice in vocal fold nodules. This study also focuses on finding the predominant clinical voice quality in the voice of patients affected by the vocal fold nodules. The quantification of pathological voice quality is carried out by finding the features which are highly correlated with perceptual voice quality.

This paper is organised as follows: Section 2 reviews the related work. Section 3 describes the database and data collection procedure. Pre-processing and feature extraction are explained in Section 4. Feature engineering and classification process is given in Section 5. The method for feature selection is stated in Section 6. Severity assessment methodology is given in Section 7. Section 8 provides the result and discussion. Finally, the conclusion is drawn in Section 9.

2 Previous work

Al-Nasheri et al. (2017a) proposed an automatic system for voice pathology assessment by extracting maximum peak value from the correlation functions as a feature to detect and classify disordered samples. English, German, and Arab databases are used in this study. Sustained phonation of vowel /a/ of normal and pathological voices are extracted for analysis. SVM is used as a classifier in this work.

Muhammad et al. (2017) developed a method to detect voice pathology using an interlaced derivative pattern (IDP), MFCC, and multidimensional voice profile (MDVP) parameters and SVM classifier. It is observed that IDP obtained better classification results for experiments conducted using three different databases, i.e., MEEI, SVD, and AVPD.

Silva et al. (2019) presented a method to assess voice pathology using acoustic features jitter, shimmer, autocorrelation, and HNR. The classification is achieved using multilayer perceptron (MLP) ANN. The sustained vowels /a/, /i/, and /u/ from the German database were analysed from healthy subjects and pathological subjects such as chronic laryngitis, and vocal cord paralysis.

Hemmerling et al. (2016) extracted a feature containing 28 coefficients extracted from time, frequency, and cepstral domain features from vowel utterances of /a/, /i/, and /u/. The dimensionality of the feature set was reduced using principal component analysis. The classification was performed using k-means clustering and the random forest classifier.

Al-Nasheri et al. (2017b) discussed the usefulness of MDVP parameters for automatic detection and classification of voice pathology. The databases of native speakers of three different languages are used in this study. Optimum results are obtained using SVM classifier for the three highest-ranked MDVP parameters arranged according to the Fisher discriminant ratio.

Benba et al. (2016) proposed a technique to segregate the patients who suffer from Parkinson's disease and the ones who suffer from other neurological disorders. The acoustic features such as MFCC, perceptual linear prediction (PLP), and RelAtive SpecTrAl PLP (RASTA-PLP) are extracted from these voice samples. An accuracy of 90% is obtained for the first 11 coefficients of the PLP using linear SVM kernels.

Muhammad et al. (2016) developed an automatic voice pathology detection system. The model used features extracted from the vocal tract area which is connected to vocal folds. The measurement of vocal tract area irregularity is done using a 24-order LP filter.

Saldanha et al. (2014) discriminated normal and pathological speech samples including samples of various disorders for sustained phonation of vowel /a/ from the MEEI database. Speech features such as MFCC and LPCC are used to characterise between healthy and pathological samples. The performance of different classifiers such as k-NN, and linear discriminant analysis (LDA) classifier is compared for the feature set.

Gidaye et al. (2020) estimated glottal velocity waveform using glottal inverse filtering to capture sub-glottal pressures. To analyse the voice of a subject with a disordered larynx various time and frequency-based glottal descriptors were estimated from the glottal volume velocity waveform. The information gain feature ranking technique is used to select the most appropriate feature with a students T-test to validate the results.

Most of the previous work is carried out using the public databases available on the website. In the proposed work an attempt is made to acoustically and perceptually analyse the speech samples of vocal nodules collected from the local hospital of Dakshina Kannada District. Spectral descriptors are used for the first time to evaluate their efficacy in characterising pathological voice. A novel function from MATLAB 2021a is used in automating the process of feature engineering. Feature engineering transforms the extracted features into most relevant features for optimising the accuracy of classification. The most relevant feature for characterising vocal nodules is estimated using the MRMR feature ranking method. A set of speech pathologists evaluated the perceptual quality of voice. In order to quantify and assess the severity level of pathological voice quality, the correlation between each feature and its corresponding perceptual quality is evaluated.

3 Database

Ethical approval for data collection was obtained from Father Muller Institutional Ethics Committee, Father Muller Research Center and Hospital, Mangalore. A detailed information sheet consisting of purpose, procedure, duration, participant selection, and confidentiality was given to participants. Participants who volunteered in this study signed the informed consent form. For the children in the age group of 10 to 18, the Child Assent Form and Parental Consent Form were given for taking the consent for data collection. The information content and consent forms were provided in both Kannada (local language) and English.

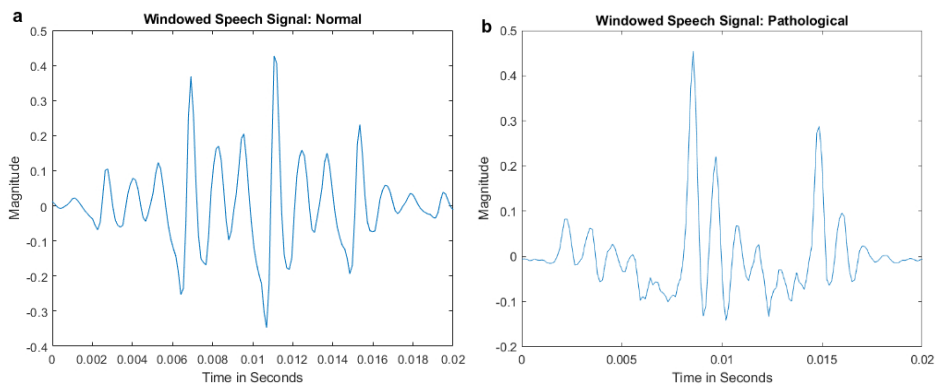
The data was collected in the Speech and Voice Laboratory at Father Muller College of Speech and Hearing, Mangalore. The laboratory is a sound-treated room that exhibits nominal ambient noise. The speech samples are picked up by a fixed microphone positioned at a distance of 15 cms from the speaker's mouth. The speech samples were recorded in the presence of the speech pathologist. During the data collection, each participant was instructed to sit straight with a straight-backed chair and asked to utter sustained phonation of vowel /a/ at their comfortable pitch and loudness. The speaking task was first demonstrated to the participant by the speech pathologist, and after a short practice session, the voice recording was done to obtain speech samples for study.

Speech corpus collected consisted of both male and female voice samples of normal and pathological categories belonging to vocal fold nodules in the age group from 10 to 70 years. The diagnosis and the confirmation of the vocal nodules was carried out by an experienced laryngologist through endoscopy procedure at rest position, before initiating medical treatment. The voice samples of the individuals with psychogenic voice disorders, phono surgery, and the individuals who have undergone radiation therapy of neck areas, were excluded from this study. Participants reported no medical history of smoking, hearing, speech, or language disorders. Participants were free from laryngeal inflammation due to cold or other allergies. The subjects who participated in this study were informed about the process and consent was taken from them. This protocol was approved by the parent institution's ethics committee.

4 Pre-processing and feature extraction

The speech signal is a highly time-varying signal. In order to extract features, the speech signal is subdivided into overlapping frames of 20–30 msec using hamming window. Figures 1(a) and 1(b) shows windowed speech segments of normal and pathological speech samples respectively for the sustained vowel /a/ for a duration of 20 msec. Normal speech samples exhibit periodic patterns with a similar pattern repeating at regular intervals (pitch periods). Whereas the pathological speech sample has noisy nature due to the fact that the presence of nodules cause irregular vibration pattern of the vocal folds.

Figure 1 Windowed speech segments for sustained vowel /a/, (a) normal speech sample
(b) pathological speech sample (see online version for colours)



The speech spectrum is one of the useful representations for finding the harmonic and non-harmonic nature of speech signals. A vowel spectra of a healthy speech sample is dominated by harmonic components whereas vowel spectra of pathological speech sample is dominated by non-harmonic/noisy components as shown in Figure 2. This acts as a motivation to extract various spectral descriptors for the analysis of pathological voice. Various cepstral and pitch perturbation features are also extracted at this stage. The speech samples are converted into feature vectors of 18 dimensions. Various features extracted in this study for the analysis of normal and pathological speech samples are described in Figure 2.

Spectral tilt is used to describe the overall slope of the power spectral density. The spectral tilt is low when the upper tone is dominant and high when the lower tone is dominant (Peeters, 2004). The pathological voice samples are expected to produce a lower spectral tilt due to the presence of higher energy at upper tones. Spectral tilt corresponding to bin k is calculated as

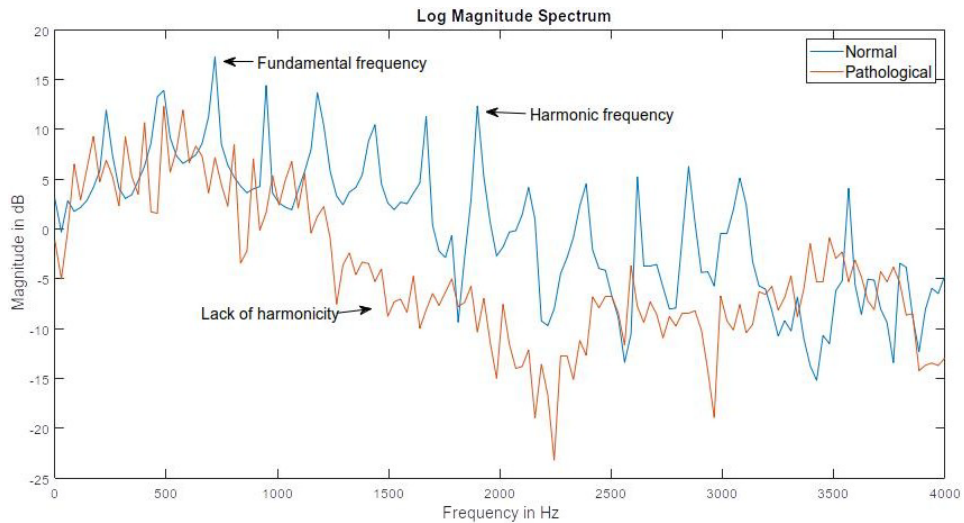
$$\text{Spectral tilt} = \frac{\sum_{k=b_{e1}}^{b_{e2}} (fr_k - \mu_1)^3 sp_k}{\mu_2^3 \sum_{k=b_{e1}}^{b_{e2}} sp_k} \quad (1)$$

where

- fr_k is the frequency

- sp_k is the spectral magnitude
- be_1 and be_2 are the edges of the band
- μ_1 is the spectral centroid
- μ_2 is the spectral spread.

Figure 2 Spectrum of speech sample for sustained phonation of /a/ (see online version for colours)



Spectral spread is the spread of the spectrum around its mean value. Spectral spread increases as the tone diverges and decreases as the tone converges. So for a pathological voice, the spectral spread is expected to be high as compared to the normal voice. Spectral spread corresponding to bin k is calculated as

$$Spectral\ spread = \sqrt{\frac{\sum_{k=be_1}^{be_2} (fr_k - \mu_1)^2 sp_k}{\sum_{k=be_1}^{be_2} sp_k}} \quad (2)$$

where

- fr_k is the frequency
- sp_k is the spectral magnitude
- be_1 and be_2 are the edges of the band
- μ_1 is the spectral centroid.

Spectral entropy is used to find the peakiness of the probability mass function. Spectral entropy is found to be useful in differentiating voiced and unvoiced sounds (Misra et al., 2004). For voiced sounds, the spectrum consists of clear peaks representing formant frequencies. Probability mass function of such peaky spectrum will have lower entropy.

Whereas the spectrum of pathological sounds will have flat distribution resulting in high entropy. Spectral entropy corresponding to bin k is given as

$$\text{Spectral entropy} = \frac{-\sum_{k=be_1}^{be_2} sp_k \log(sp_k)}{\log(be_2 - be_1)} \quad (3)$$

where

- fr_k is the frequency
- sp_k is the spectral magnitude
- be_1 and be_2 are the edges of the band.

The spectral kurtosis is a measure of the flatness of the spectrum around its centroid (Peeters, 2004). As the noise is increased due to the presence of pathology, the spectral peaks become less dominant hence decreasing the kurtosis. Spectral kurtosis corresponding to bin k is calculated as

$$\text{Spectral kurtosis} = \frac{\sum_{k=be_1}^{be_2} (fr_k - \mu_1)^4 sp_k}{\mu_2^4 \sum_{k=be_1}^{be_2} sp_k} \quad (4)$$

where

- fr_k is the frequency
- sp_k is the spectral magnitude
- be_1 and be_2 are the edges of the band
- μ_1 is the spectral centroid
- μ_2 is the spectral spread.

The spectral crest is an indication of the peakiness of the spectrum. A periodic signal is expected to have a higher spectral crest, whereas a pathological signal will have a lower spectral crest due to its noisy nature (Peeters, 2004). The spectral crest corresponding to bin k is given as

$$\text{Spectral crest} = \frac{\max(sp_{k[be_1, be_2]})}{\frac{1}{be_2 - be_1} \sum_{k=be_1}^{be_2} sp_k} \quad (5)$$

where

- sp_k is the spectral magnitude
- be_1 and be_2 are the edges of the band.

Spectral flatness is defined as the ratio of the geometric mean to the arithmetic mean of a power spectrum (Johnston, 1988). A higher value of spectral flatness indicates that the

signal is noisy. A healthy speech sample is expected to have a lower value of spectral flatness. The spectral flatness corresponding to bin k is given as

$$\text{Spectral flatness} = \frac{\left(\prod_{k=be_1}^{be_2} sp_k \right)^{\frac{1}{be_2-be_1}}}{\frac{1}{be_2-be_1} \sum_{k=be_1}^{be_2} sp_k} \quad (6)$$

where

- sp_k is the spectral magnitude
- be_1 and be_2 are the edges of the band.

Speech signal exhibits low energy at high frequencies. This property of the energy-frequency relationship can be well characterised by spectral roll-off. It is also a measure of spectral shape. Spectral roll-off represents the frequency below which 95% of the magnitude distribution of the spectrum is concentrated in the spectrum. It can be used to distinguish between the voiced and unvoiced speech (Scheirer and Slaney, 1997), hence can be effectively used for pathological voice analysis. The spectral roll-off point corresponding to bin k is given as

$$\text{Spectral roll-off point} = i, \text{ such that } \sum_{k=be_1}^i |sp_k| = k \sum_{k=be_1}^{be_2} sp_k \quad (7)$$

where

- sp_k is the spectral magnitude
- be_1 and be_2 are the edges of the band
- k is the energy threshold.

The spectral centroid indicates at which frequency the energy of a spectrum is centered. A higher value of spectral centroid represents a peaky spectrum with dominant formants. It is a measure of shape and position of the spectrum. It is commonly used in music analysis (Peeters, 2004) and voiced/unvoiced speech detection (Raimy and Cairns, 2015). Spectral centroid corresponding to bin k is calculated as

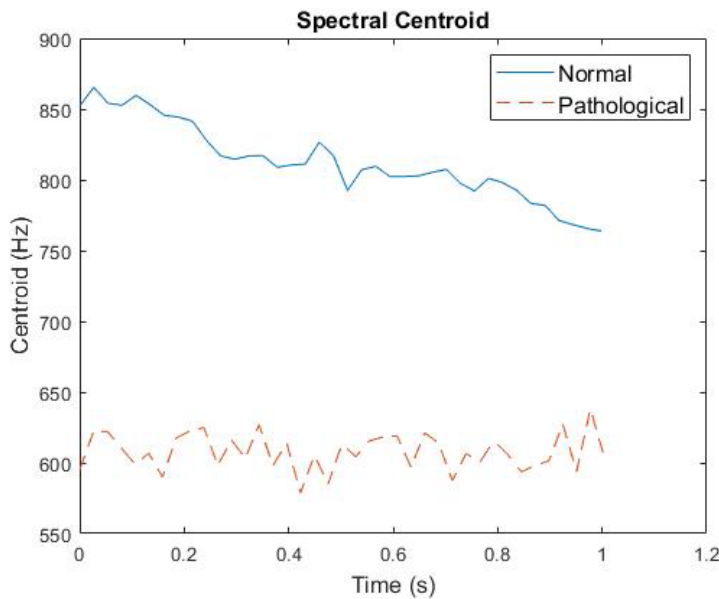
$$\text{Spectral centroid} = \frac{\sum_{k=be_1}^{be_2} fr_k sp_k}{\sum_{k=be_1}^{be_2} sp_k} \quad (8)$$

where

- fr_k is the frequency
- sp_k is the spectral magnitude
- be_1 and be_2 are the edges of the band.

Sustained phonation of vowel /a/ results in voiced speech. Voiced speech is generated by an exciting vocal tract with periodic glottal pulses from vibrating vocal folds due to which the voiced spectrum consists of harmonics of glottal source and peaks in the spectral envelope representing formants. Unvoiced speech is generated by exciting vocal tract by random noise. Presence of vocal fold nodules in the larynx results in irregular vibration of vocal folds resulting into a speech signal which exhibits the properties of unvoiced speech. A spectrum of such signal lacks harmonics of the source and dominant peaks representing resonating frequencies of the vocal tract. The wide difference in the spectral centroid values of the speech sample for normal and pathological sample in Figure 3 shows that spectral centroid can be used effectively to discriminate normal and vocal nodule speech sample.

Figure 3 Spectral centroid of normal and pathological speech sample (see online version for colours)



The spectral slope describes how rapidly the amplitudes of successive frequency components decrease as frequency increases (Hansen and Patil, 2007; Lerch, 2012). The spectral slope values are obtained by curve fitting envelopes of the spectrum. The resonant frequency of the vocal tract is related to the slope, it is noticeable when the energy at lower formants is more than the energy at higher formants. Pathological speech sample tends to have weak formants in the spectrum of speech. This property makes it useful for pathological speech analysis in vocal nodules. The spectral slope corresponding to bin k is given as

$$\text{Spectral slope} = \frac{\sum_{k=be_1}^{be_2} (fr_k - \mu_f)(sp_k - \mu_s)}{\sum_{k=be_1}^{be_2} (fr_k - \mu_f)^2} \quad (9)$$

where

- fr_k is the frequency
- μ_f is the mean frequency
- sp_k is the spectral magnitude
- μ_s is the mean spectral value.

All the spectral features used in this work are extracted from Hamming windowed speech segments of 50 ms duration with 50% overlap. One sided power spectrum is used for the extraction of spectral features. The frequency range from 0 to 4,000 Hz (Nyquist frequency) is used for each frame for the calculation of spectral features used in this study.

Jitter is the parameter of the frequency variation between successive cycles of the speech signal (Saldanha et al., 2023). Jitter is impacted primarily by the absence of control of vibration of the cords. The voice of the patients with vocal fold nodules often has a higher percentage of jitter. There are two important variants of jitter, namely jitter absolute which is the average absolute difference between two consecutive periods, and jitter local which is the ratio of average absolute difference and average period (Teixeira and Gonsalves, 2016).

Shimmer is the variation in amplitude of the consecutive periods. Shimmer local is obtained by taking the ratio of average absolute difference between the amplitudes of the consecutive periods and the average amplitude (Teixeira and Gonsalves, 2016). The pathological speech samples are expected to exhibit a high value of shimmer than the normal speech samples because of irregular vocal fold vibration due to the presence of lesions in the larynx. Shimmer local and shimmer in decibel parameters are used in this work. Jitter and shimmer variants are calculated for every 10 msec frame.

HNR and noise to harmonic ratio (NHR) features have been used to assess vocal fold behaviour due to their ability to quantify noise in the signal. The correlation-based method is used to find the HNR and NHR (Boersma, 1993). A lower HNR and a higher NHR indicate dysphonia.

The harmonic ratio indicates the ratio of energy in the harmonic portion of the signal to the total energy of the signal (Kim et al., 2006). The harmonic structure depends on the waveform produced by the vibrating vocal folds. The presence of non-harmonic components in the speech spectrum increases the degree of roughness, harshness, or hoarseness in the perceived signal. Non-harmonic components are introduced due to irregular vocal fold vibration in vocal nodules.

Cepstral peak prominence (CPP) is obtained by taking the inverse fast Fourier transform (FFT) of the log power spectrum of the voice signal. It is the difference between the cepstral peak and the regression line drawn directly below the peak (Fraile and Godino-Llorente, 2014). Thus, signal which are periodic in nature have a better harmonic structure, obtaining a higher value of CPP. Smoothing of cepstrum before extracting cepstral peak results in cepstral peak prominence smoothed (CPPS) (Castellana et al., 2017).

5 Feature engineering and classification

Feature engineering is a process of transforming raw data into a form that better represents the predictive model. Feature engineering is performed on the extracted features using various functions available at statistics and machine learning toolbox in MATLAB R2021a. Table 1 shows the list of engineered features obtained from the original set of features using feature engineering. The feature engineering method carries out different transformations of the original features. All transformed and original features are standardised using different data standardisation methods. The function also uses in-built MRMR feature selection method in forming the engineered feature set. The engineered feature set includes subset of both the transformed features and original standardised features based of their relevance for classification. The training data consisting of 18 features are transformed into a new feature set with ten features. The validation and test data sets are also transformed into engineered features using the feature transformer object created while performing feature engineering.

Table 1 Description of the engineered features

<i>Engineered features</i>	<i>Input variables</i>	<i>Transformations</i>
Harmonic ratio.*spectral centroid	Harmonic ratio, spectral centroid	Harmonic ratio.*spectral centroid standardisation with z-score (mean = 494.5297, std = 280.4022)
CPP. * spectral centroid	CPP, spectral centroid	CPP.*spectral centroid standardisation with z-score (mean = 14,374.1274, std = 9250.3853)
NHR	NHR	Equi-probable binning (number of bins = 15)
Jitter local	Jitter local	Equi-probable binning (number of bins = 16)
Spectral RollOff Point./NHR	Spectral roll-off point, NHR	Spectral RollOff Point./NHR standardisation with z-score (mean = 64,582.5159, std = 71,189.7185)
CPPS.*spectral centroid	CPPS, spectral centroid	CPPS.*spectral centroid standardisation with z-score (mean = 3,840.5709, std = 2,367.0477)
CPP	CPP	Equi-probable binning (number of bins = 16)
Spectral entropy	Spectral entropy	Equi-probable binning (number of bins = 16)
Spectral spread	Spectral spread	Equi-probable binning (number of bins = 16)
Spectral centroid-spectral spread	Spectral centroid, spectral spread	Spectral centroid – spectral spread standardisation with z-score (mean = 3,840.5709, std = 2,367.0477)

Note: std: standard deviation.

An optimisable decision tree that maximises the hyper-parameters is used for classification. The training model takes 30 iterations. Five-fold cross-validation is used on the training dataset. The maximum number of splits and split criteria are chosen to be hyper-parameters. The training model tries a different combination of values

for hyper-parameters to minimise the model classification error. The model chooses maximum number of splits to be 7 and maximum deviance reduction as split criteria. These are the optimised hyper parameters obtaining minimum classification error when the model completes the optimisation process.

6 Feature selection

Feature selection (Zemmal et al., 2018) is widely used to find the feature with maximum discriminative power. The process of feature selection prevents over-fitting by reducing the dimensionality of the data and improving the predictive power of the model. In the contest of this study, feature selection is used to find the most relevant feature capable of characterising the voice affected by vocal fold nodules. A rank-based MRMR feature selection method (Ding and Peng, 2005) has gained more popularity after the publication by Uber engineers (Zhao et al., 2019). The algorithm is found to be the most effective among all other conventional feature selection methods. A majority of feature selection methods consider linear relationship between the features and response classes. Although these methods ignore the interdependency among the features. MRMR feature selection method used in this work considers linear dependency between features and their response classes, as well as interdependency between the features for classification. It maximises the mutual information between the features and target classes, and minimises the mutual information between the features. The algorithm maximises the relevance (V_s) of the feature set to a response variable and minimises the redundancy (W_s) of the feature set.

The relevance and redundancy of the feature set are given as

$$V_s = \frac{1}{|S|} \sum_{x \in S} I(x, y) \quad (10)$$

$$W_s = \frac{1}{|S|^2} \sum_{x, z \in S} I(x, z) \quad (11)$$

$|S|$ is number of features in S .

The mutual information I of the discrete random variable X and Z is defined as

$$I(X, Z) = \sum_{i, j} P(X = x_i, Z = z_j) \times \log \frac{P(X = x_i, Z = z_j)}{P(X = x_i)P(Z = z_j)} \quad (12)$$

$$MIQ_x = \frac{V_x}{W_x} \quad (13)$$

where V_x and W_x are the relevance and redundancy of a feature x , respectively.

$$V_x = I(x, y) \quad (14)$$

$$W_x = \frac{1}{|S|} \sum_{z \in S} I(x, z) \quad (15)$$

The algorithm returns features ordered by their importance and their corresponding scores. The feature with the maximum score is most important. The rank is calculated based on the maximum value of V_x and MIQ_x .

7 Severity assessment

Severity is assessed using correlation between features and perceptual voice quality. A set of five specialists from the speech pathology department of Father Muller College of Speech and Hearing evaluated the perceptual quality of voice and its severity. Specialists rated the voice quality as healthy, mild, moderate, and severe on a scale of 0 to 4. As per the ratings obtained from the specialists, the predominant voice quality for vocal fold nodules is found to be hoarse. Pearson correlation coefficient between hoarseness and each of the features is evaluated.

8 Results and discussion

The classification model uses a total of 231 samples, in which 70 samples belong to the normal category and 161 samples belong to vocal fold nodules. Samples in the train and test data set were divided into a 70:30 ratio. Figures 4(a) and 4(b) shows the minimum classification error plot of the original set of features and engineered features for the train validation dataset respectively. For each iteration, the model tries to update the plot with the least amount of classification error that has been observed up to that iteration. The red square in the plot depicts the point of optimised hyper-parameters obtained when the model completes the optimisation process. It can be noticed that point of optimised hyper-parameters is achieved much faster in the case of engineered features than the original set of features. Figures 5(a) and 5(b) shows the train validation receiver operating characteristic curve for original and feature engineered dataset respectively. The improvement in area under curve (AUC) of receiver operating characteristic curve of feature engineered dataset shows that there is improvement in the true positive rate of the classifier. Figures 6(a) and 6(b) shows the confusion matrix of optimised decision tree classifier for original and feature engineered test dataset respectively. In Figure 6, it is observed that the number of samples incorrectly classified is reduced from 4 to 3 for class 0 (pathological) and from 10 to 8 for class 1 (normal). Table 2 shows the classification results of the optimised decision tree classifier for the original and feature engineered dataset. For the feature engineered dataset, there is an improvement in all performance measures evaluated, with classification accuracy of 84.2%.

Figure 4 Minimum classification error plot for, (a) original dataset (b) engineered features (see online version for colours)

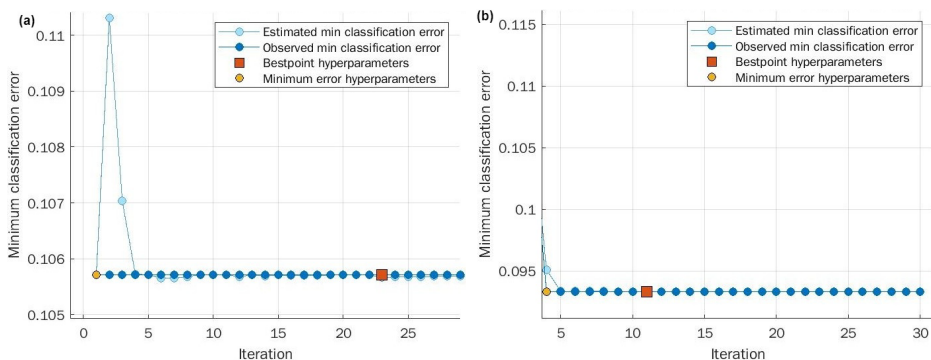


Figure 5 Train validation ROC for, (a) original dataset (b) feature engineered dataset (see online version for colours)

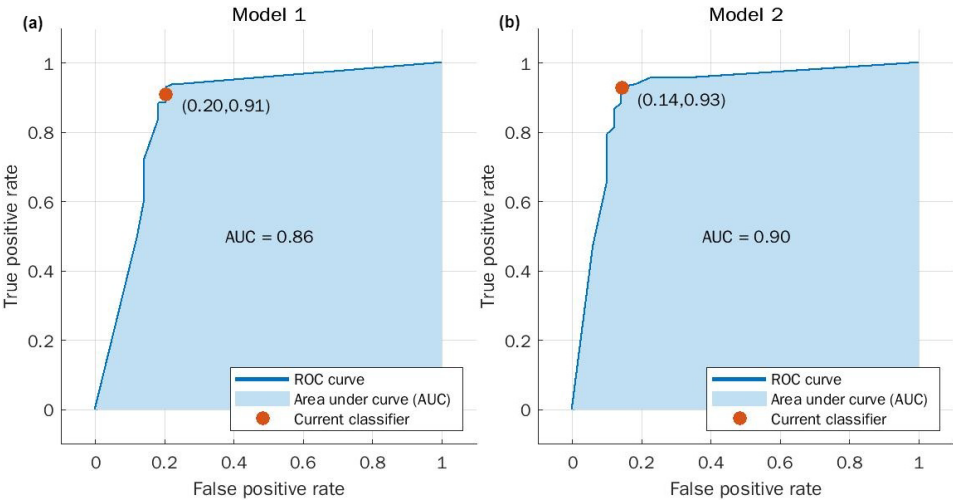


Figure 6 Confusion matrix of the optimised decision tree classifier for, (a) original test dataset (b) feature engineered test dataset (see online version for colours)

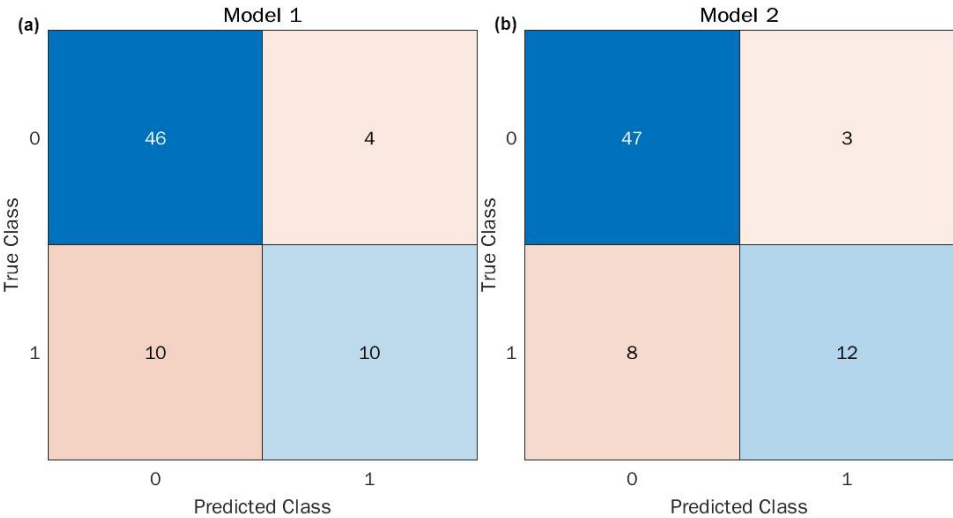


Table 3 shows the classification results of a set of optimised classifiers namely optimised decision tree, bagged trees, linear and quadratic SVM, Naive Bayes, and ANN. It is observed that the decision tree classifier outperforms all the other classifiers. The classifiers obtained significantly lower values of specificity due to the biasing of classifiers as a result of imbalance in the number of samples in the classification categories. Figure 7 shows the feature ranking based on the MRMR algorithm for each of the predictors (features) used in this work. Spectral centroid feature obtained the highest rank with significantly higher value of predictor importance score as compared to all other features used in this study.

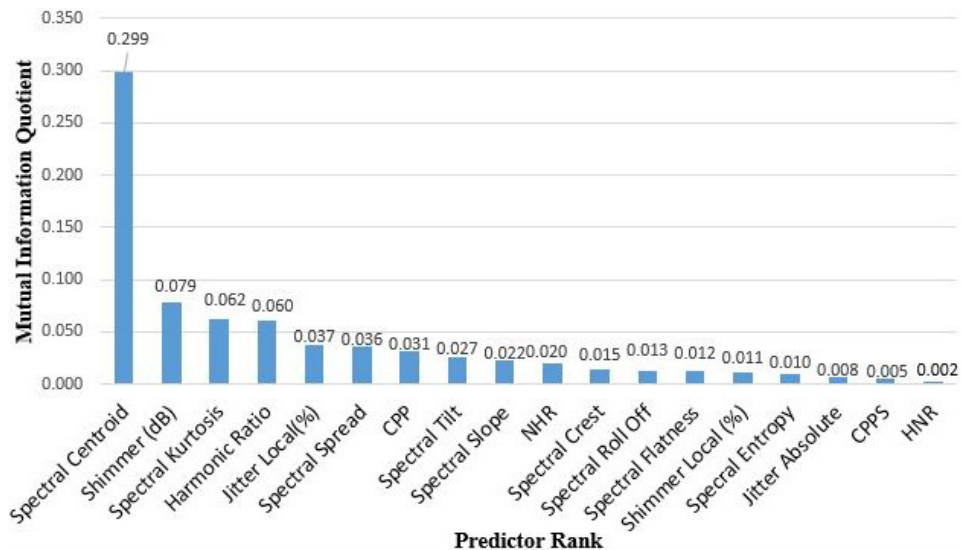
Table 2 Classification results of optimised decision tree classifier on original and feature engineered dataset

	<i>Sensitivity (%)</i>	<i>Specificity (%)</i>	<i>MCC</i>	<i>F1-score</i>	<i>Positive predictive value (PPV)</i>	<i>Negative predictive value (NPV)</i>	<i>Accuracy (%)</i>
Original features	92	50	0.56	0.86	0.82	0.71	80
Engineered features	94	60	0.65	0.89	0.85	0.80	84.2

Notes: MCC: Matthews correlation coefficient; F1-score: Fisher-1 score.

Table 3 Classification results of optimised classifiers on the feature engineered test dataset

	<i>Sensitivity (%)</i>	<i>Specificity (%)</i>	<i>MCC</i>	<i>F1-score</i>	<i>Positive predictive value (PPV)</i>	<i>Negative predictive value (NPV)</i>	<i>Accuracy (%)</i>
Decision tree	94	60	0.60	0.89	0.85	0.80	84.2
Bagged trees	98	45	0.58	0.89	0.82	0.90	82.8
Linear SVM	92	55	0.60	0.87	0.84	0.73	81.4
Quadratic SVM	82.4	50	0.58	0.87	0.82	0.77	81.4
Naive Bayes	84	60	0.56	0.84	0.84	0.60	77
Neural network	96	50	0.60	0.88	0.82	0.83	82.8

Figure 7 Feature ranking using MRMR algorithm (see online version for colours)

Feature ranking obtained for MRMR algorithm is compared with feature ranking obtained for correlation feature selection (CFS) method. It is observed that the spectral

centroid feature obtained first rank among all the features in both MRMR and CFS methods. Table 4 shows the feature ranking in case of MRMR and CFS algorithms. Spectral centroid obtaining the topmost rank proves that spectral centroid is highly correlated with the target classes. Figure 8 shows the discriminative capability of spectral centroid features between normal and vocal fold nodule pathology classes.

Figure 8 Scatter plot of spectral centroid between normal and pathological classes (see online version for colours)

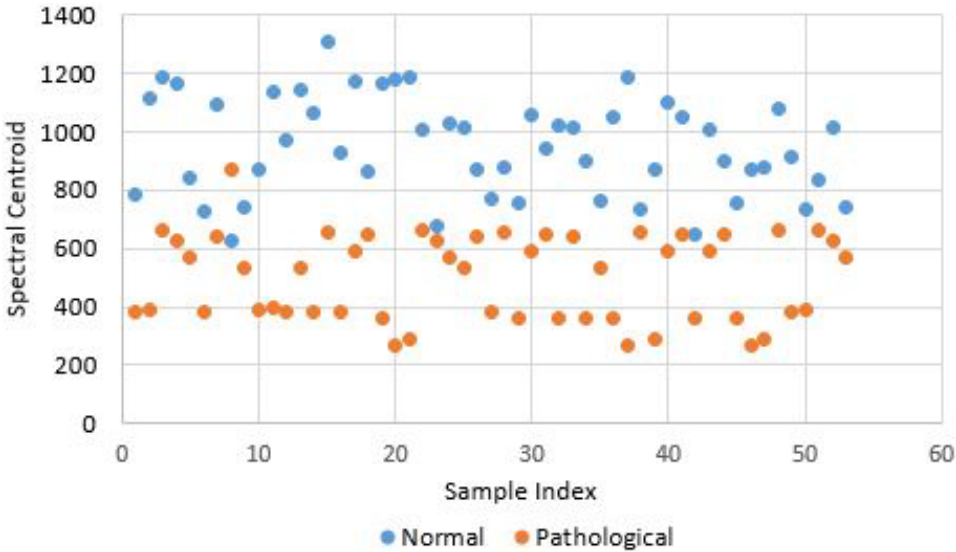
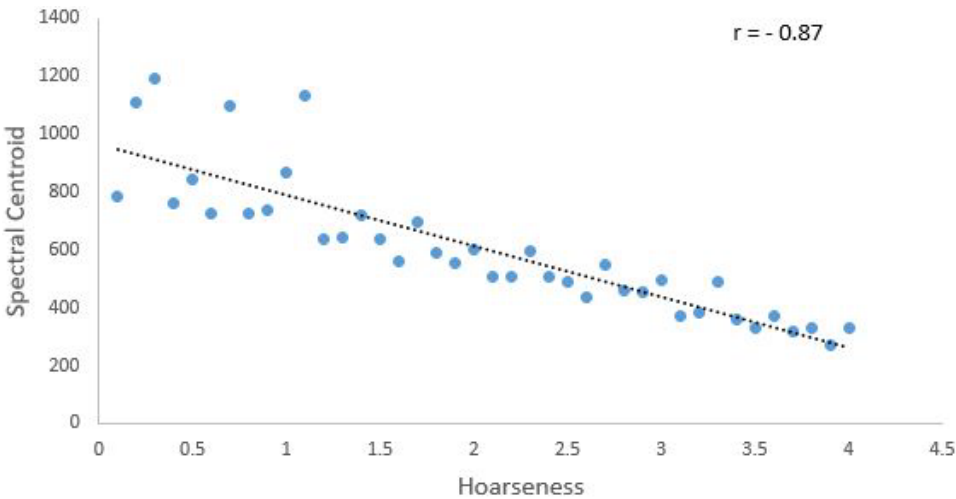


Figure 9 Scatter plot of hoarseness and spectral centroid (see online version for colours)



A correlation coefficient between all extracted features and the level of severity in the hoarseness of pathological voice is computed. Among all spectral descriptive features, the spectral centroid feature is found to be strongly negatively correlated with a

correlation coefficient of -0.87 . Figure 9 shows the physical correlation between the spectral centroid and pathological voice in vocal fold nodules. Figures 10 to 13 shows the scatter plots of hoarseness with harmonic ratio, shimmer (dB), shimmer local (%), and HNR features. These features obtained a significant amount of correlation with hoarseness as compared to all other features used in this study. Table 5 shows correlation coefficient for shimmer (dB), shimmer local(%), HNR, harmonic ratio, and spectral centroid.

Table 4 Comparison of feature ranking in MRMR and CFS algorithms

Rank	Feature ranking in MRMR	Feature ranking in CFS
1	Spectral centroid	Spectral centroid
2	Shimmer (dB)	Harmonic ratio
3	Spectral kurtosis	Spectral roll off point
4	Harmonic ratio	Shimmer (dB)
5	Jitter local (%)	Spectral entropy
6	Spectral spread	Spectral crest
7	CPP	Shimmer local (%)
8	Spectral tilt	Spectral slope
9	Spectral slope	Spectral tilt
10	NHR	CPP
11	Spectral crest	Spectral kurtosis
12	Spectral roll off point	Spectral spread
13	Spectral flatness	NHR
14	Shimmer local (%)	Jitter local (%)
15	Spectral entropy	Jitter absolute
16	Jitter absolute	HNR
17	CPPS	Spectral flatness
18	HNR	CPPS

Table 5 Pearson correlation coefficient for features

Features	Pearson correlation coefficient (r)
Spectral centroid	-0.84
Shimmer local (%)	0.7
Shimmer (dB)	-0.73
Harmonic ratio	-0.8
HNR	-0.79

This paper investigates the features relevant for automating the assessment of vocal fold nodules using the voice samples. The analysis is carried out using sustained phonation of vowel since discriminative information between normal and disordered samples is expected to be more significant than the running speech samples. A set of features including pitch perturbation features such as jitter, shimmer, and cepstral features such as CPP and CPPS along with various spectral descriptive features are extracted to evaluate their efficacy in characterising voice affected by vocal fold nodules. Feature engineering is applied to the data set using the most recently introduced function *gencfeatures* from MATLAB R2021a. The main goal of the feature engineering function is to transform

the features into a form that is more relevant for classification purposes. The decision tree classifier performance is evaluated on both original and feature engineered data set. It was also observed that there is an increase in classification accuracy by 4.2% with feature engineered dataset.

Figure 10 Scatter plot of hoarseness and harmonic ratio (see online version for colours)

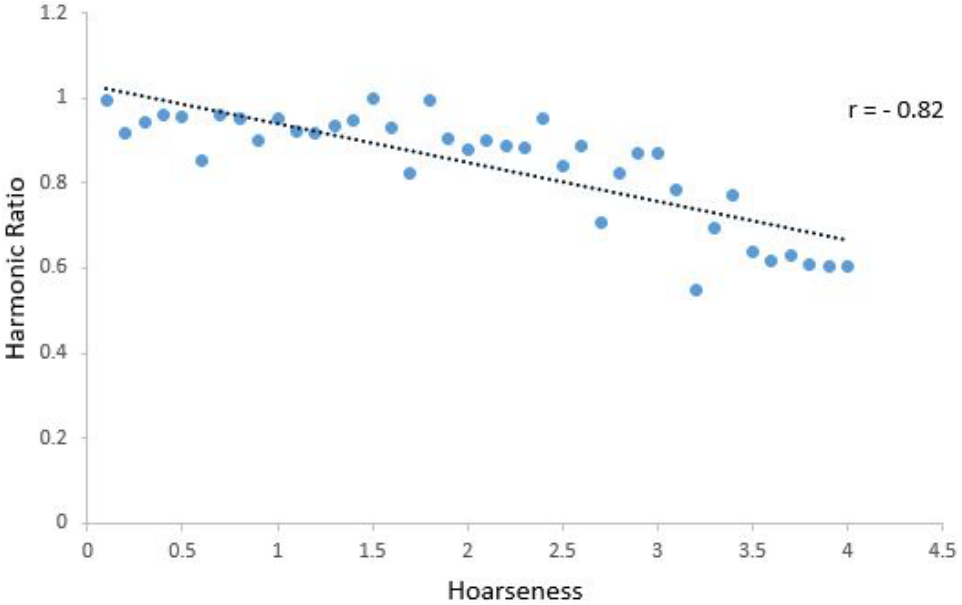
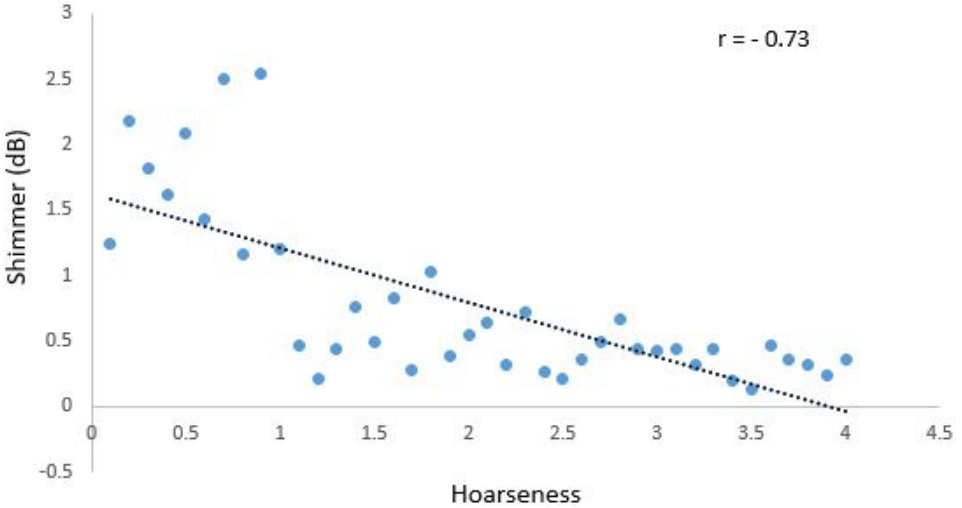


Figure 11 Scatter plot of hoarseness and shimmer (dB) (see online version for colours)



Feature engineering plays a significant role in increasing classification accuracy. The feature engineering tool performs feature transformation and feature selection. In the process of feature transformation, the algorithm generates different combinations of

original features. All original features and transformed features are standardised using z-score and equiprobable binning methods. The z-score method scales the data to have a mean of 0 and a standard deviation of 1, which makes that data more interpretable by limiting the impact of outliers and skewness on the data. Furthermore, z-scores can boost the performance and accuracy of a variety of algorithms. In equi-probable binning, the algorithm splits the variable values into the number of bins of equal probability, to enhance the interpretability of the data analysis. An MRMR feature ranking algorithm applied to the standardised data further reduces the redundancy between features thereby increasing the relevance of features to the classification categories and obtaining a smaller and more accurate dataset for classification. The decision tree classifier obtained the highest classification accuracy compared to all the other classifiers used in this study. Decision trees are the most powerful predictive models with the highest discriminative power. The accuracy of SVM classifiers depends on the choice of the kernel function. Linear or nonlinear kernel functions must be selected based on the distribution of data. The decision tree does not require any prior knowledge of the distribution of data. Decision tree classifiers have high precision with low computation complexity. They are simple to implement and interpret. In the case of bagged trees, several such decision trees are combined and an average of all the predictions from different trees are used. Bagged trees are useful to prevent over-fitting which may happen in the case of decision tree classifiers. ANN has high computational complexity as compared to decision trees. ANN can outperform the decision tree if the size of the dataset is large. If the requirement is low computational complexity and high accuracy decision tree can be preferred over ANN. Naive Bayes classifier is a generative algorithm based on a probabilistic approach. Naive Bayes classifier is more accurate with a large dataset. It assumes that the features are conditionally independent while this assumption is often violated in practice (Rish, 2001). The classification results may vary with the size of the dataset. Hence depending on the requirement, the appropriate choice of the classifier can be made. As the speech data is collected from the native Kannada speakers of Dakshina Kannada district of Karnataka, India, the effect of variations in the speech features on the classification, due to the difference in native language is ruled out.

An MRMR feature selection method which ranks the features based on their relevance is used to find the most relevant feature to characterise the dysphonic voice in vocal fold nodules. The performance of MRMR algorithm is compared with CFS method. The spectral centroid feature evaluated to be the feature with highest correlation ranking. According to the functioning of the algorithms, CFS algorithm selects the feature which is highly correlated with target class, whereas MRMR algorithm not only selects the feature which is highly correlated with the target class but also feature having minimum mutual information among all features in the feature set. Spectral centroid feature is highly correlated with the classification categories but highly uncorrelated with all the features in the feature set which is evident from the feature rankings obtained from MRMR and CFS algorithms. Spectral centroid describes the center of gravity, since center of gravity is related to shape and the position of peaks in the spectrum (Hosseinzadeh and Krishnan, 2007), it implies that spectral centroid can track the variations in pitch and harmonic structure in the speech spectrum of normal and vocal nodule samples. As spectral centroid carries maximum information about the energy distribution in the speech spectrum it can be considered to be more relevant than any other spectral features considered in this study.

Figure 12 Scatter plot of hoarseness and shimmer local (see online version for colours)

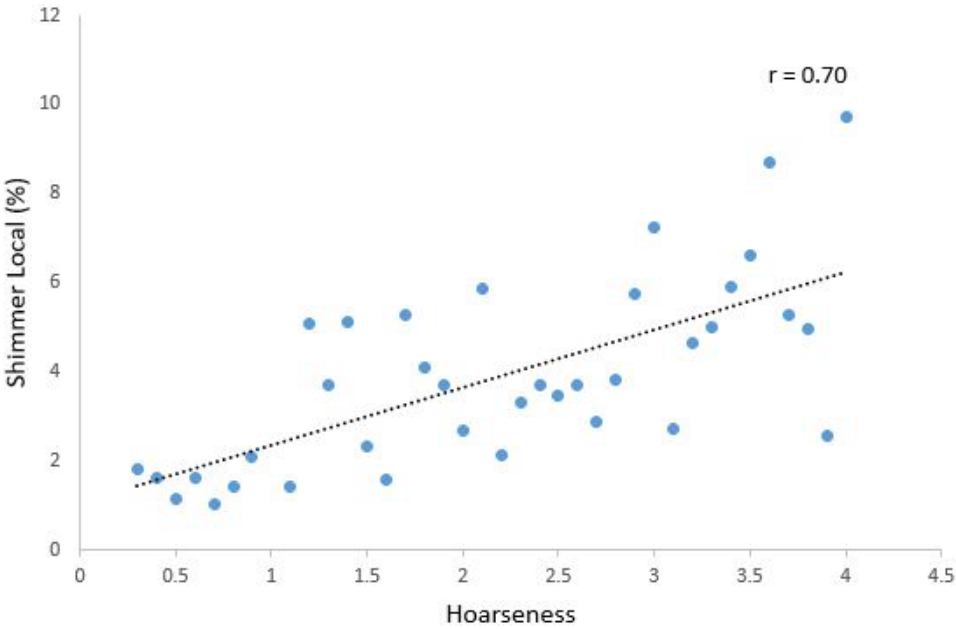
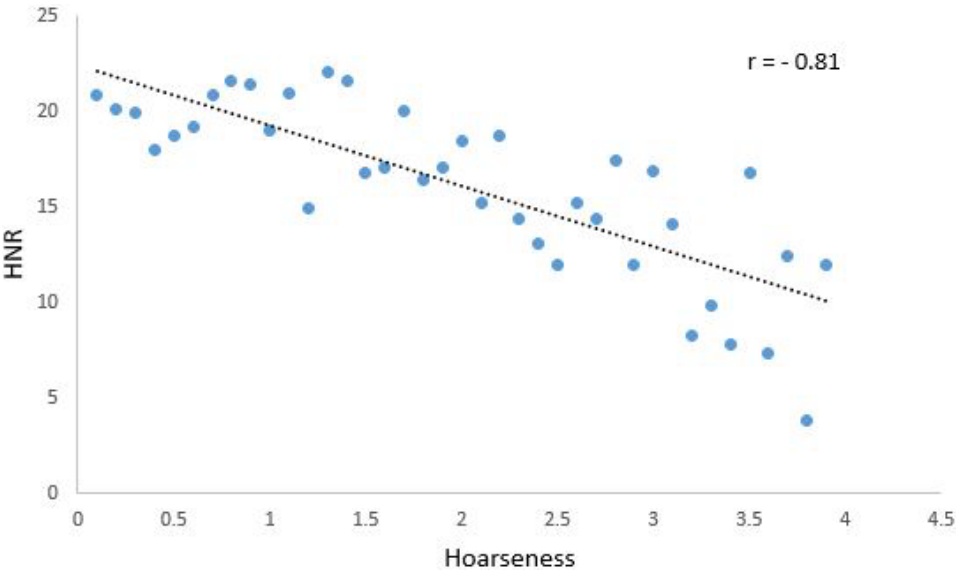


Figure 13 Scatter plot of hoarseness and HNR (see online version for colours)



Five specialists evaluated the voice samples perceptually and rated the severity of pathological voices by Harsh, Breathy, and Hoarse voice types on the scale from 0 to 4. On an average rating obtained from the specialist, it was found that 95% of the voice samples were evaluated to be hoarse. The statistical analysis is carried out to find the feature suitable for modelling the hoarse voice quality of dysphonic voice.

Researchers have found different features useful for objective assessment of degree of hoarseness in the voice. HNR is a useful indicator of hoarseness in the voice (Yumoto et al., 1982). Dysphonic severity of rough (harsh) voices can be modelled by HNR, breathy voice can be modelled by jitter, shimmer, and HNR (Martin et al., 1995). Shimmer is a quantitative descriptor of severity level of hoarseness in the dysphonic voice (Lopes et al., 2014). None of the earlier studies (Al-Nasheri et al., 2017a, 2017b; Muhammad et al., 2017, 2016; Silva et al., 2019; Hemmerling et al., 2016; Benba et al., 2016) have reported the detection of vocal nodules. One of the study (Szkłanny and Wrzeciono, 2019) has addressed the detection of vocal nodules using genetic algorithm. The study used sustained phonation of vowel /a/ from normal and vocal nodules disordered subjects from a private Polish database of children. The database consisted of voice recordings of children in the age group of 5 to 14. Genetic algorithm is used for classification obtaining the training accuracy of 92%. The authors also compared the results obtained with deep neural network classifier which gave 81.81% train accuracy. Table 6 shows the comparison of sustained phonation of vowel /a/ samples of normal and vocal fold nodule speech samples between some of the public databases namely, MEEI, SVD, and the private database used in this study. The features such as glottal-to-noise excitation ratio, jitter, and shimmer are extracted from public databases. The classification between normal and vocal fold nodule samples is carried out using a random forest classifier, implemented on extracted features (Madruga et al., 2020). The results of the optimised decision tree classifier on feature-engineered data extracted from the private database used in this work are tabulated in the last column of the table for performance comparison. It can be observed that the results of the proposed method are comparable with the existing work in the literature.

Table 6 Performance comparison of standard public database with the private database

<i>Database</i>	<i>MEEI</i>	<i>SVD</i>	<i>Private database</i>
Accuracy (%)	95	62	84.2
Specificity (%)	87	55	60
Sensitivity (%)	98	69	94

The current work aims at using optimisable classifiers which optimises hyper parameters as part training process. The hyper parameters which minimises the misclassification error are used for classification. Decision tree outperforms all the other classifiers used in this work with maximum training accuracy as 90.7% for original features. The aim of this work is not to improve the accuracy of existing model, instead to find the feature that characterises voice affected by vocal nodules disorder. Spectral centroid feature not only obtained maximum accuracy with decision tree classifier but also found to be highly correlated with hoarse voice quality resulting due to the presence of nodules. In our study we have explored the usage of a latest feature namely the spectral centroid in the assessment of pathological voice. The centroid tends to have strong negative correlation with hoarseness. The harmonic ratio, HNR, and variants of shimmer are found to be less correlated as compared with the spectral centroid. Hence, spectral centroid can be effective in modelling dysphonic voice which is evident from the feature ranking obtained by both MRMR and CFS feature ranking methods.

9 Conclusions

In this work, automatic detection of vocal fold nodules from voice samples of Dakshina Kannada District of Karnataka State, India is carried out. A set of spectral descriptive measures, cepstral features, and pitch perturbation measures were extracted. Feature engineering is performed on the original dataset using MATLAB 2021a. The decision tree classifier performance is evaluated on the feature engineered and non-feature engineered dataset. An improvement of 4.2% accuracy is obtained with feature engineered dataset. The performance of the decision tree classifier is compared with a set of optimisable classifiers such as bagged trees, Naive Bayes, ANN, linear and quadratic SVM. The decision tree classifier obtained maximum classification accuracy compared to the other classifiers on feature engineered dataset. However, the specificity of the classifiers can be improved by using a balanced dataset. In order to investigate the features relevant to characterise the pathological voice, features are ranked on the basis of their relevance using the MRMR algorithm. The spectral centroid feature is found to be most accurate, and relevant for characterising voice affected by vocal fold nodules. Perceptual analysis of pathological voices proved that the predominant voice quality for vocal fold nodules is hoarse. Correlation analysis proved that spectral centroid is highly correlated (negatively) with hoarseness in the voice. Hence, spectral centroid can be effectively used for objective assessment of the level of severity of pathological voice quality in the vocal nodules. Future work includes implementing deep learning, transfer learning, and hybrid learning models for the detection of vocal fold nodules using a larger data set. This work discusses a methodology for assessing the dysphonia in vocal fold nodules that is not region or language specific. Although, the current study is carried out using the database of native Kannada speakers of Dakshina Kannada District, it is believed that the results can also be generalised to other languages.

Acknowledgements

We would like to thank speech-language pathologists from the Department of Speech and Language Pathology, College of Speech and Hearing, Father Muller Charitable Institutions, Mangalore, for their proceeded cooperation in assisting us with assessing the sound recordings of the patients with vocal nodules disorder used in this study.

References

- Al-Nasheri, A., Muhammad, G., Alsulaiman, M. and Ali, Z. (2017) 'Investigation of voice pathology detection and classification on different frequency regions using correlation functions', *Journal of Voice*, Vol. 31, No. 1, pp.3–15.
- Al-Nasheri, A., Muhammad, G., Alsulaiman, M., Ali, Z., Mesallam, T.A., Farahat, M. and Bencherif, M.A. (2017) 'An investigation of multidimensional voice program parameters in three different databases for voice pathology detection and classification', *Journal of Voice*, Vol. 31, No. 1, p.113-e9.
- Bajpai, C., Ramnarayan, S. and Naik, K.J. (2023) 'Deep learning model for plant-leaf disease detection in precision agriculture', *International Journal of Intelligent Systems Technologies and Applications*, Vol. 21, No. 1, pp.72–91.

- Benba, A., Jilbab, A. and Hammouch, A. (2016) 'Discriminating between patients with Parkinson's and neurological diseases using cepstral analysis', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 24, No. 10, pp.1100–1108.
- Boersma, P. (1993) 'Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound', in *Proceedings of the Institute of Phonetic Sciences*, March, Vol. 17, No. 1193, pp.97–110.
- Castellana, A., Carullo, A., Corbellini, S., Astolfi, A., Bisetti, M.S. and Colombini, J. (2017) 'Cepstral peak prominence smoothed distribution as discriminator of vocal health in sustained vowel', in *2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, Turin, Italy, May, pp.1–6.
- Ding, C. and Peng, H. (2005) 'Minimum redundancy feature selection from microarray gene expression data', *Journal of Bioinformatics and Computational Biology*, Vol. 3, No. 2, pp.185–205.
- Dodderi, T., Philip, N.E. and Mutum, K. (2018) 'Prevalence of voice disorders in the Department of Speech-Language Pathology of a tertiary care hospital of Mangaluru: a retrospective study of 11 years', *Journal of Health and Allied Sciences NU*, Vol. 8, No. 3, pp.12–16.
- Fraile, R. and Godino-Llorente, J.I. (2014) 'Cepstral peak prominence: a comprehensive analysis', *Biomedical Signal Processing and Control*, Vol. 14, No. 2, pp.42–54.
- Gidaye, G., Nirmal, J., Ezzine, K., Shrivasa, A. and Frikha, M. (2020) 'Application of glottal flow descriptors for pathological voice diagnosis', *International Journal of Speech Technology*, Vol. 23, No. 1, pp.205–222.
- Hansen, J.H. and Patil, S. (2007) 'Speech under stress: analysis, modeling and recognition', in *Speaker Classification I*, pp.108–137, Springer, Berlin, Heidelberg.
- Harar, P., Galaz, Z., Alonso-Hernandez, J.B., Mekyska, J., Burget, R. and Smekal, Z. (2018) 'Towards robust voice pathology detection', *Neural Computing and Applications*, pp.1–11.
- Hemmerling, D., Skalski, A. and Gajda, J. (2016) 'Voice data mining for laryngeal pathology assessment', *Computers in Biology and Medicine*, Vol. 69, No. 2, pp.270–276.
- Hosseinzadeh, D. and Krishnan, S. (2007) 'Combining vocal source and MFCC features for enhanced speaker recognition performance using GMMs', in *2007 IEEE 9th Workshop on Multimedia Signal Processing*, IEEE, October, pp.365–368.
- Johnston, J.D. (1988) 'Transform coding of audio signals using perceptual noise criteria', *IEEE Journal on Selected Areas in Communications*, Vol. 6, No. 2, pp.314–323.
- Kim, H.G., Moreau, N. and Sikora, T. (2006) *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*, No. 3, John Wiley & Sons.
- Lerch, A. (2012) *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*, No. 1, Wiley-IEEE Press.
- Lopes, L.W., Cavalcante, D.P. and Costa, P.O.D. (2014) 'Severity of voice disorders: integration of perceptual and acoustic data in dysphonic patients', in *CoDAS*, Sociedade Brasileira de Fonoaudiologia, September, Vol. 26, pp.382–388.
- Madruca, M., Campos-Roca, Y. and Perez, C.J. (2020) 'Multicondition training for noise-robust detection of benign vocal fold lesions from recorded speech', *IEEE Access*, Vol. 9, No. 1, pp.1707–1722.
- Martin, D., Fitch, J. and Wolfe, V. (1995) 'Pathologic voice type and the acoustic prediction of severity', *Journal of Speech, Language, and Hearing Research*, Vol. 38, No. 4, pp.765–771.
- Misra, H., Ikbali, S., Bourlard, H. and Hermansky, H. (2004) 'Spectral entropy based feature for robust ASR', in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, QC, Canada, May, Vol. 1, pp.1–193.
- Muhammad, G., Altuwaijri, G., Alsulaiman, M., Ali, Z., Mesallam, T.A., Farahat, M. and Al-Nasheri, A. (2016) 'Automatic voice pathology detection and classification using vocal tract area irregularity', *Biocybernetics and Biomedical Engineering*, Vol. 36, No. 2, pp.309–317.

- Muhammad, G., Alsulaiman, M., Ali, Z., Mesallam, T.A., Farahat, M., Malki, K.H. and Bencherif, M.A. (2017) 'Voice pathology detection using interlaced derivative pattern on glottal source excitation', *Biomedical Signal Processing and Control*, Vol. 31, No. 2, pp.156–164.
- Naunheim, M.R. and Carroll, T.L. (2017) 'Benign vocal fold lesions: update on nomenclature, cause, diagnosis, and treatment', *Current Opinion in Otolaryngology & Head and Neck Surgery*, Vol. 25, No. 6, pp.453–458.
- Peeters, G. (2004) 'A large set of audio features for sound description (similarity and classification) in the CUIDADO project', *CUIDADO 1st Project Report*, Vol. 54, pp.1–25.
- Raimy, E. and Cairns, C.E. (2015) *The Segment in Phonetics and Phonology*, No. 1, John Wiley & Sons.
- Rish, I. (2001) 'An empirical study of the Naive Bayes classifier', in *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, August, Vol. 3, No. 22, pp.41–46.
- Saldanha, J.C., Ananthakrishna, T. and Pinto, R. (2014) 'Vocal fold pathology assessment using mel-frequency cepstral coefficients and linear predictive cepstral coefficients features', *Journal of Medical Imaging and Health Informatics*, Vol. 4, No. 2, pp.168–173.
- Saldanha, J.C., Suvarna, M., Satish, D. and Santhmayor, C. (2023) 'Jitter as a quantitative indicator of dysphonia in Parkinson's disease', *International Journal of Intelligent Systems Technologies and Applications*, Vol. 21, No. 2, pp.199–228.
- Scheirer, E. and Slaney, M. (1997) 'Construction and evaluation of a robust multifeature speech/music discriminator', in *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE, April, Vol. 2, pp.1331–1334.
- Schindler, A., Mozzanica, F., Ginocchio, D., Maruzzi, P., Atac, M. and Ottaviani, F. (2012) 'Vocal improvement after voice therapy in the treatment of benign vocal fold lesions', *Acta Otorhinolaryngologica Italica*, Vol. 32, No. 5, p.304.
- Silva, L., Hermsdorf, J., Guedes, V., Teixeira, F., Fernandes, J., Bispo, B. and Teixeira, J.P. (2019) 'Outliers treatment to improve the recognition of voice pathologies', *Procedia Computer Science*, Vol. 64, pp.678–685.
- Szklanny, K. and Wrzeciono, P. (2019) 'The application of a genetic algorithm in the noninvasive assessment of vocal nodules in children', *IEEE Access*, Vol. 7, No. 2, pp.44966–44976.
- Teixeira, J.P. and Gonsalves, A. (2016) 'Algorithm for jitter and shimmer measurement in pathologic voices', *Procedia Computer Science*, Vol. 100, pp.271–279.
- Yumoto, E., Gould, W.J. and Baer, T. (1982) 'Harmonics to noise ratio as an index of the degree of hoarseness', *The Journal of the Acoustical Society of America*, Vol. 71, No. 6, pp.1544–1550.
- Zemmal, N., Nabih A., Mokhtar S., Djamel Z., Soraya C., Nilanjan D. and Amira S.A. (2018) 'Robust feature selection algorithm based on transductive SVM wrapper and genetic algorithm: application on computer-aided glaucoma classification', *International Journal of Intelligent Systems Technologies and Applications*, Vol. 17, No. 3, pp.310–346.
- Zhao, Z., Anand, R. and Wang, M. (2019) 'Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform', in *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, Washington, DC, USA, October, pp.442–452.