IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

VOICE DISORDER CLASSIFICATION USING MACHINE LEARNING TECHNIQUES

[1]Mrs. A Shifa Sultana, [2]Mrs. Reshma P
[1,2]Assistant Professor
[1,2]Electrical and Electronics Engineering,
[1,2]The National Institute of Engineering, Mysore, India-570008

Abstract : This paper is basically based on evaluation of multiple speech disorders. The methodology discussed here is based on (ANN) Artificial Neural Networks with Machine Learning Algorithm. Other classifiers which are considered are Decision Tree and Random Forest which will give a better idea with the classification accuracy for voice impairment. In this paper few voice disorders considered are Parkinson Disease (PD), Laryngeal Pathologies (LP), Cleft lip and palate (CLP) and normal. The acoustic based method is generally used in the evaluation of voice disorders, the outcome depends on the listener and the expertise to get a reproducible valuation. The dataset taken of patients speaks the vowels basically. Vowels are considered because they give certain variations in the voice which makes the prediction of disorder easy for the machine when trained and tested. Once the data is taken first, they are pre-processed i.e., the noise which are available in the samples are removed where there is very minimum loss of data. Once this is done feature extraction of the samples are carried out under frequency and time domain. Frequency domain extraction are done mainly for the voice or word extraction from the sample using the fundamental frequency. Mel -frequency cepstral coefficients (MFCC), Linear predictive coding (LPC), Wavelet Packet Decomposition (WPD) are the extraction done in Frequency domain whereas Cepstral Analysis (CA), Jitter, Shimmer, Intensity, Harmonicity, Pulse & Pitch and Energy & Entropy are carried under Time domain. These methods utilized ranges its accuracy from 82% to 88% in the automatic detection of voice impairments. The study and results obtained from this suggests that neural networks algorithms give better accuracy then other classifiers utilized. Also, its necessary to choose appropriate set of features to model the voice pathologies.

***Index Terms* - Voice Disorders, Parkinson's disease (PD), Pathologies (LP), Cleft lip and palate (CLP), Mel - frequency cepstral coefficients (MFCC), Linear predictive coding (LPC), Wavelet Packet Decomposition (WPD), Cepstral Analysis (CA), Jitter, Shimmer, Intensity, Harmonicity, Pulse & Pitch, Energy & Entropy, Artificial neural network, Machine Learning Techniques.**

INTRODUCTION

Individuals with risk of pathological voice problems are drastically increasing day by day. Roughly 25% of the world population whose professions require them to speak excessively louder than the normal intensity [1] suffer from these kinds of problems, for e.g. singers, teachers, actors, lawyers, auctioneers, instructors and supervisors are all considered to work in professions that make heavy command on their voice. As a corollary, working on the digital processing of speech signals has been found to provide a skewed noninvasive methods such as direct inspection using endoscopic instruments [2] like Laryngoscopy, glottography, stroboscopy, but these techniques are very expensive in time and human assets for this reason we need a vigorous and accurate method that enables us to gauge the disorder type and its brutality to evaluate the patient condition. A voice disorder is categorized by the abnormal making and/or absences of vocal quality, pitch, loudness, resonance, and/or duration, which is inappropriate for an individual's age and/or sex. The pathological struggle like vocal nodules, edema, vocal fold paralysis, and neurological problems like brain tumour, lesions, neural degeneration, brain injury may affect the speech producing section of the brain. Thus the voice contains the concealed information about disorders of nervous system. Disorders reduce the effectiveness of movement as well as the ability to speak clearly. Tumour, brain injury, growth of lesions, neural degeneration in the cells of brain which controls the speech communication will affect the speech production. A voice disorder is a problem involving abnormal pitch, loudness or quality of the sound produced by the larynx, more generally known as the voice box. Almost every disorder may present in more than one symptom and one cannot correlate one single sign with specific voice disorder. For example, hoarseness, increased vocal effort or limitation in pitch and loudness may be a sign of any number of disorders. Severity of the voice symptoms differs according to the disorder and the individual based on the age, severity, duration and effect on the capacity of a person . Voice disorders might exist in both adults and as well as in children. Understanding voice disorders by knowing how normal voice is produced and the role of the voice box and its parts play in speaking helps us in understanding the voice disorders. Voice Disorders and their different origins: Neurological, Functional and Laryngeal [4]

A voice is declared "disordered" when the vocal quality of an individual is altered/changed in such a way that it is thought to be abnormal to the percipient. The commencement and development of these disorders can be "sudden" or "slow". The voice disorders are caused due to change in mass of vocal chords, injury to vocal chords, neurological problems which can affect the speech producing organs. The voice disorders are mainly classified as neurological and pathological voice disorders. In neurological voice disorder the voice producing center may get affected or nerve supply for the voice producing organ may be damaged. In pathological voice disorder the voice producing organs may get damaged or infected. When vocal folds do not work harmoniously, there exist impairment in general quality of the voice. Neurological voice disorders are difficult to detect using skewed technique based on medical instrument, for this instance researchers are mainly focusing more on the implementation of various types of feature extraction techniques[3] from speech signal, which are Mel - frequency cepstral coefficients (MFCC), Linear predictive coding (LPC), Wavelet Packet Decomposition (WPD) , Cepstral Analysis (CA), Jitter, Shimmer, Intensity, Harmonicity, Pulse & Pitch and Energy & Entropy for the diagnosis of these diseases.

Neurological Voice Disorders: Neurologic voice disorders symptoms go down into two categories. Dysarthria is difficulty in setting up words or in enunciating words correctly, making a person tough to recognize. A mechanical speech disorder is when the parts of the mouth that form words go wrong. Typical speech production requires control and coordination of the muscles [6] of the voice box, throat, sense of taste, jaw, tongue and lips. Neurological problems are because of irregularities of the cerebrum as well as the nerves of the body. These irregularities bring about hindered control of the muscles of the voice box, throat, sense of taste, jaw, tongue or lips, causing an assortment of voice or potentially discourse issues. The neurological issue prompts shortcoming of voice box muscles, sense of taste muscles, jaw, lip and tongue muscles. Parkinson's disease is a dynamic sensory system issue that influences motion and movement. Symptoms start bit by bit, some of the time beginning with a scarcely perceptible tremor in only one hand. Tremors are normal, yet the turmoil additionally generally causes solidness or easing back of development. In the beginning periods of Parkinson's infection, your face may show next to zero appearance to any expression. Your arms may not sway when you stroll. Your talk may turn out to be delicate or slurred. Parkinson's disease side effects intensify as your condition advances after some time. Despite the fact that Parkinson's disease can't be reassured, drugs may altogether improve your side effects. At times, your primary care physician may propose medical procedure to direct certain locales of your cerebrum and improve your side effects.

Functional Voice Disorders: Functional voice disorders are related with issues controlling various appendages or potentially muscles engaged with the speech establishment process. Congenital fissure(Cleft lip and palate) and sense of taste is one of the most unmistakable craniofacial contortion and it creates diverse utilitarian issues [8,9] in the vocal tract. Congenital fissure and congenital fissure are openings or parts in the upper lip, the top of the mouth (sense of taste) or both. Congenital fissure and congenital fissure result when facial structures that are creating in an unborn child don't close totally. Congenital fissure and congenital fissure are among the most well-known birth surrenders. They most normally happen as disconnected birth absconds but at the same time are related with many acquired hereditary conditions or disorder. Having an infant brought into the world with a split can be upsetting, yet congenital fissure and congenital fissure can be remedied. In many children, a progression of medical procedures can re-establish typical capacity and accomplish an increasingly ordinary appearance with insignificant scarring

Laryngeal Voice Disorders: Laryngeal disorders cause dysphonia, which is disability of the voice. A productive change in the voice requires observation [10] of the vocal cords, including their adaptability. Despite the fact that the voice changes with propelling age, turning out to be hoarse and aperiodic, intense or conspicuous changes in the elder ought not be attempted to come about because of aging, and assessment is compulsory. The voice have to be evaluated and recorded, especially if surgeries are arranged. Assessment of the larynx incorporates outside review and palpation of the neck and inward representation of the epiglottis, false cords, true cords, arytenoids, pyriform sinuses, and subglottic section beneath the cords. Inside representation is cultivated by either aberrant mirror assessment or direct adaptable fiberoptic laryngoscopy in an outpatient setting with a topical sedation. Inflexible laryngoscopy with the patient under general sedation gives the most exhaustive assessment of the vocal lines, permitting, visualization of the under surfaces, assessment of inactive portability when immobilized by either paralysis or fixation, and biopsy

When relaxed, the vocal cords typically structure a V-formed opening that permits air to go completely through to the trachea. The cords open the airway during inspiration and close the airway route during gulping or speech. At the point when a mirror is held in the rear of a patient's mouth, the vocal cords can frequently be seen and checked for disorders, for example, contact ulcers, polyps, knobs, loss of motion, and malignant growth. Loss of motion may influence (one- sided) or both vocal cords.
The main objective of this work is to investigate the parameters of feature extracted from the dataset and to build and train a classification model, for this purpose, the classification algorithms used are Artificial neural network (ANN), random forest classifier (RFC) and decision tree (DT).

RELATED WORK

The following literatures are the works relevant to characterizing the voice disorder, Here we provide a brief overview of some recent findings related to our research topic.

M.S Hossain et al [12] have said Speech or, in general, the voice signal is used in several kinds of application ranging from emotion recognition to patient healthcare state recognition . Speech processing is a very active area of research. They are many contributions that focus on different aspects of speech processing, from feature extraction to decision support systems based on speech analysis. A. Al-Nasheri et al [1] have concentrated on developing feature extraction for the detection and classification of voice pathology by investigating different frequency bands through which we found a variation in the accuracies in the same database. M. Dahmani et al [2] have experimented determination of different types of vocal folds pathologies as it is present in articulatory details of vocal folds in frequency and time domain. T. B. Ijitona et al [3] have adopted an extended speech feature to classify disordered speech from healthy speech using artificial neural network. J. R. Orozco-Arroyave et al [4] have adopted features to discriminate different aspects of voice through noise content, spectral- cepstral modelling, and nonlinear behavior over functional, neurological and laryngeal disorders. L. Verde et al [5] have used multimedia services and application in healthcare sector for monitoring and detection of diseases and have used several machine learning techniques to classify the disorder. S. Firdos et al [6] have adopted the MFCC feature extraction on Multi-level SVM is used to classify the normal, neurological and pathological voices which will help speech therapists to detect the pathology at initial stage. K, Uma et al [7] have experimented on various acoustic features which are extracted by time domain and frequency domain techniques then later done by fusing different classifiers which are fed with features extracted from different domains. E. B. Hook et al [8] tell us the detail report on the Congenital malformations worldwide on the international monitoring system. D. Sell et al [9] gives the case study of clept lip and palate care in united kingdom. J. Hillenbrand [10] have conducted on the purpose of the study to extend these results to speakers with laryngeal pathologies and to conduct tests using connected speech in addition to sustained vowels. Crovato et al [11] have depicted dysphonic voice classification system using the wavelet packet transform and the best basis algorithm (BBA) as dimensionality reductor and 06 artificial neural networks (ANN) acting as specialist systems.

MATERIAL AND METHOD

A. Database

The database is an important factor for the development of our system which allows the evaluation of the algorithm and the comparison of the result obtain. Our work is evaluated using Massachusetts Eye and Ear Infirmary (MEEI), the link to download the dataset is https://www.masseyeandear.org/.

The dataset collected in this paper include 3 voice disorders those are **,** Parkinson's disease (PD), Pathologies (LP), Cleft lip and palate (CLP) and one normal voice signal. The data collected is in Spanish language and in .wav form. We have total of 339 samples of normal and disordered voice signal.

B. Feature Extraction used for Classification

Feature extraction is an essential mission that allows an enhancement of the examination and categorization. Instead of giving directly the whole data samples, we are extracting the features so that it is easy for the machine to train it, doing so we can save memory, processing time and accuracy as well. Feature Extraction is done in both Time and Frequency domain, in Frequency domain we include Mel-Frequency Cepstral Coefficients (MFCC), Linear Predictive Analysis (LPC) and Wavelet Packet Decomposition (WPD), while the feature extracted in Time domain include Cepstral Analysis (CA), Jitter & Shimmer, Intensity, Harmonicity, Pulse & Pitch, Energy & Entropy. The Frequency Domain refers to the analytic space in which mathematical functions or signals are conveyed in terms of frequency, rather than time. For example, where a time- domain graph may display changes over time, a frequency-domain graph displays how much of the signal is present among each given frequency band. It is possible, however, to convert the information from a time-domain to a frequency-domain. An example of such transformation is a Fourier transform. The Fourier transform converts the time function into a set of sine waves that represent different frequencies. The frequency-domain representation of a signal is known as the spectrum of frequency components. The total number of features included in our work are 7722.

B.1. Frequency Domain Features

The Frequency domain features are Mel-Frequency Cepstral Coefficients (MFCC), Linear Predictive Analysis (LPC) and Wavelet Packet Decomposition (WPD)

B.1.1 Mel-Frequency Cepstral Coefficients (MFCC): These coefficients attempt to investigate the vocal tract without assistance of the vocal folds that are harmed because of voice pathologies. In this work, the investigations were led utilizing 39 MFCC coefficients. MFCC values are not extremely tough within the sight of added core disorder, thus it is entirely expected to standardize their qualities in discourse acknowledgment frameworks to decrease the impact of roar. A few specialists propose adjustments to the fundamental MFCC calculation [5] to improve energy for example, by raising the log- mel-amplitudes to a reasonable force (around 2 or 3) preceding taking the DCT (Discrete Cosine Transform), which lessens the impact of low-vitality parts.

B.1.2 Linear Predictive Analysis (LPC): To reduce computational intensive process of traversing and computational complexity and finding the source and system components from time domain itself, the Linear Prediction analysis [2] is created. The essential target of LPC is representing the spectral envelope of a digital signal of speech in compressed form using the information of a linear predictive model and to compute the LP coefficients by reducing the prediction error. The well-known technique for calculating the LP coefficients by least squares auto correlation method. This is accomplished by limiting the total prediction error. Doing so we can get great quality speech at a low bit rate and provide high accurate rating of speech variable parameters.

B.1.3 Wavelet Packet Decomposition(WPD) : Wavelet Transform is a arithmetical function, which are minuscule waves located in different time, it is obtained using scaling and translation of the scaling functions and wavelet functions. Thus wavelet is generalized in both time and frequency domain [11].

Wavelet does it work by cutting the data into different frequency component and then study each component with a resolution matched with the scale. They are infinite set of basic functions in wavelet, we are importing this using open source wavelet transform software for python programming language, using Daubecies wave basic function

B.2. Time Domain Extraction

Time domain alludes to examination of mathematical functions, physical signal or time series of data with respect to time. An oscilloscope is an apparatus frequently used to envision real-world signals in the time domain space. A time-domain depicts how a signal changes with time, whereas a frequency-domain tells how much of the signal lies within each given frequency band over a range of frequencies.

The time domain features [5] are Cepstral Analysis (CA), Jitter, Shimmer, Intensity, Harmonicity, Pulse & Pitch and Energy & Entropy

B.2.1 Cepstral Analysis (CA): Speech is formulated by excitation of source and vocal tract system components. In order to dissect and form the excitation and system components of the speech separately and also use that in variety of speech processing applications, these two segments have to be separated from the speech. The goal of cepstral analysis is to isolate the speech into its source and system components without any a earlier knowledge about source and / or system.

B.2.2 Jitter & Shimmer: The arrangement of jitter and shimmer boundaries and the Harmonic-to-Noise Ratio (HNR) [5] is introduced underneath down . The qualities for Jitter can be calculated in different parameters, such as absolute, relative, relative average perturbation (rap) and the period perturbation quotient (ppq). Jitter depicts the uncertainty of the oscillating prototype of the vocal folds, evaluating the cycle-to-cycle diversity of fundamental frequency

Shimmer (dB) indicates the unpredictability of the oscillating prototype of the vocal folds, evaluating the changeability of the peak-to-peak amplitude in decibels.

In our work, we have included the values for Jitter can be calculated in diverse parameters, such as absolute, relative, relative average perturbation (rap) and the period perturbation quotient (ppq5). Jitter absolute is the cycle-to-cycle alteration of fundamental frequency, i.e. the average absolute disparity between consecutive periods, articulated as:

$$\text{Jitter}=\frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i-1}|$$

Where Ti is the drawn out as glottal period lengths and N is the number of expressed glottal periods. Relative Jitter or local Jitter is the average absolute difference between consecutive periods, divided by the average period. It is expressed as a percentage:

$$\text{Jitter(relative)}=\frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i-1}|\frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_i-T_{i-1}|}{\frac{1}{N}\sum_{i=1}^{N}T_i}\times 100$$

Shimmer (dB) is extricated as the unevenness of the peak-to-peak amplitude in decibels, i.e. the standard absolute base-10 logarithm of the variation between the amplitude of consecutive periods, multiplied by 20.

$$\text{Shdb}=\frac{1}{N-1}\sum_{i=1}^{N-1}|20 * log\frac{(A_{i+1})}{Ai}|$$

Where Ai is the peak-to-peak amplitude data value and N is the number of extracted fundamental frequency periods.

B.2.3Intensity: Intensity or vocal intensity of the speech signal alludes to the garishness effect of speech signal. Vocal intensity is associated to the subglottis force of the airflow, which relies on the tension and the vibrations of the vocal folds. A miniature number of vibrations in the vocal folds make quitier voice as compared to the large number of vibrations of the folds. Mathematically by physical means the vocal intensity can be expressed as sound intensity level (SIL) or sound pressure level (SPL). SIL or SPL is calculated in dBs. SIL essentially tells how much louder a certain sound is as compared to the standard (soft) reference vocal intensity, of 10–12 watt/m2.

$$\text{SIL}=10\log \text{I}/I_o\text{dB}$$

where the standard intensity value and sound intensity can also be expressed in terms of SPL also. SPL=10log

P/P_o dB

Here Po is the standard pressure level and is having the value of 0.00002 Pascal. SIL and SPL tells about the same point of acoustic energy and can be used interchangeably based on the need.

B.2.4 Harmonicity: The harmonicity is a proportion of the wholeness of a harmonic formation. This paper presents a new harmonic structure that measures the extend of the regular harmonicity to a set of harmonicities. They are articulated in terms of the grid harmonicity, the temporal harmonicity, the segment-spectral harmonicity, and the segmental harmonicity. The grid harmonicity measures the wholeness of individual harmonics in every frame. The grid harmonicities in a frame are summed up to form a temporal harmonicity for signifying the strength of harmonicity. The segment-spectral harmonicity, calculated by totalled specific grid harmonicity over a section, which in turn evaluates the integrity of single separate harmonics across a segment. The segmental harmonicity conveys the total strength of harmonic structure within a segment. This set of harmonicities is accessible for a logical analysis of the harmonic structure and successful in several speech processing work. HNR is 10*log10(99/1) = 20 dB. A HNR of 0 dB means that there is equal amount of energy in the harmonics and in the noise.

B.2.5 Pulse & Pitch: Pulse &Pitch in speech, are the relative highness or lowness of a tone as perceived by the ear, which depends on the number of vibrations per second produced by the vocal cords. Pitch is the key acoustic compare of tone and modulation

B.2.6 Energy & Entropy: Energy function is used to locate approximately the time at which voiced speech become unvoiced speech and vice versa, and for high quality speech the energy can be used to distinguish speech from silence. Entropy is a method of time series analysis which is stable with respect to non-linear distortion

C. Classification

In order to make an exhaustive comparison, we have chosen different machine learning algorithms, the techniques used are Artificial Neural Network (ANN), Random Forest & Decision Tree.

C.1 Artificial Neural Network (ANN): The ANN is used to for classifying the disordered speech [3]. One of the commonly used machine learning methods is the neural network. This classification technique is robust and it combines pattern recognition with acoustic phonetics methods. A multilayer neural network with many hidden layer is used for classification. In our project work we have consider are 4 input, 512 hidden layers and 1 output layer, with the batch size of 32 samples at a time to train it. In this work we have used MLPC classifier, with logistic activation function.

C.2 Decision Tree: Decision tree builds classification or regression models in the outline of a tree formation. It disintegrate down the dataset into deflate and deflate subsets while at the same time an linked decision tree is incrementally obtained. The finishing result is a tree with decision nodes and leaf nodes. In this work we have imported linear model and tree library using python software, and we have set the maximum depth of decision tree as 500.

C.3 Random Forest: Random forests or random decision forests is an ensemble supervised learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of an individual. Random forest is used to merge many decision trees into a single model. separately, predictions prepared by decision trees (or humans) might not be accurate, but when done collective together, the predictions will be nearer to the mark on average. In this work we have imported linear model and tree library using python software, and we have set the random forest estimator as 500.

RESULTS

In the proposed research there are four voice disorders that have been considered i.e.,

- Laryngeal
- Cleft Lip
- Normal
- Parkinson

These datasets are downloaded and are trained in a manner where the output is designed to display in classes i.e., [0], [1], [2], [3] respectively. Once the trained samples are executed or run, it displays the feature extracted.
Then a pattern is followed by the machine i.e. first it shows the features then its corresponding label value also the sampling frequency and then number of features. Once after the machine ends with reading the folder, it displays the classes where in brackets it reads total 271 samples and feature length will be 7722.

Then the samples starts to iterate where this is the phase where training process starts and reduces the errors in the samples. This iteration is continued till the minimum value which means the samples are trained till it becomes zero theoretically. But in practical the machine goes till minimum. Once the iteration is stopped the accuracy is displayed with the Confusion Matrix. The confusion matrix displayed actually conveys the data of total samples taken from each class and displays the number of true and false samples from the dataset. These values may vary. Once after this according to the iterations, learning curve i.e., learning rate is shown in graph format as shown below. This curve shows the error reduction which plots the Learning rate v/s iterations.

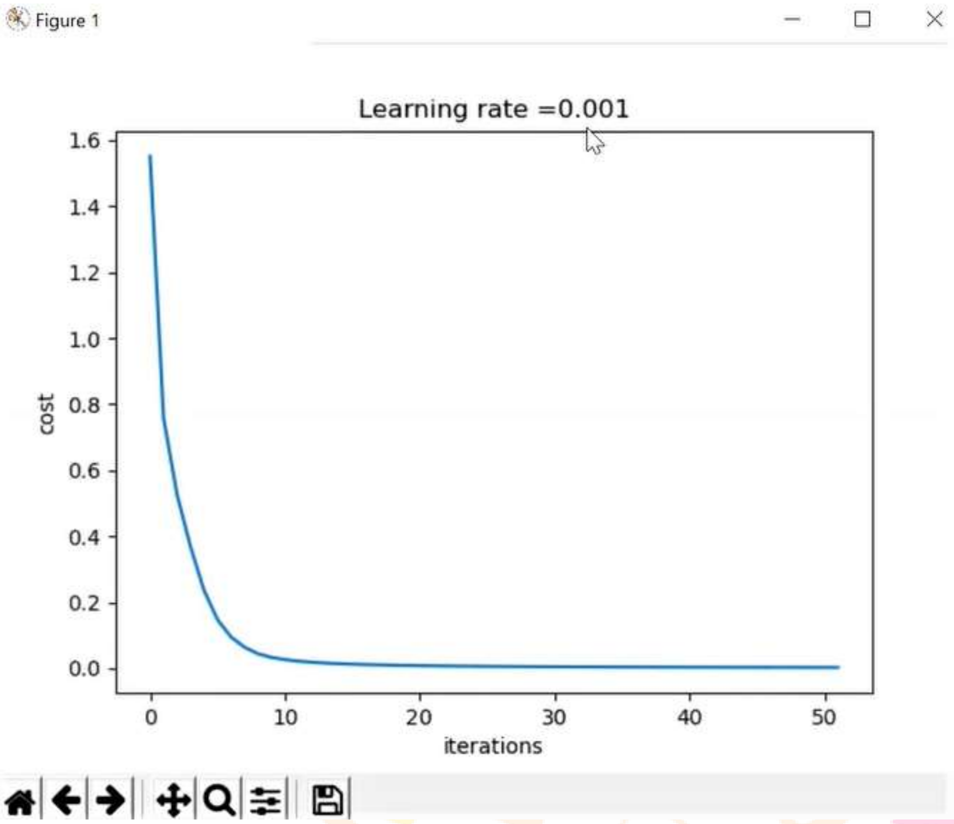

Figure 1: Shows the error reduction which plots the Learning rate v/s iterations.

Testing of samples

Once after the training, coming to testing of samples is successfully done where the sample given as input will be displayed with which voice disorder it belongs to and also its class number will be seen. Then later it shows the result for prediction for Laryngeal, functional and Parkinson before which the selection of sample is opted.

The table below shows the accuracies of different classifiers implemented in the proposed system.

CLASSIFIER	ACCURACY
ANN	88.2352
Decision Tree	38.2352
Random Forest	64.7058

The performance of the selected machine learning classification techniques was evaluated in terms of accuracy, sensibility, specificity and region of curve area by using the measurement value of True positive(TP), True Negative(TN), False positive(FP), False negative(FN). Feature selection and extraction is an important crucial task in our study to choose to test machine learning classification technique over the overall database chosen by selecting and calculating features and further classification performance is performed.

Conclusion

Four groups of features recitation different aspects of voices have been studied. The ability of each group of features to categorize between pathological and healthy voices is tested in a databases which contain recordings of patients with several pathologies including laryngeal (dysphonia due to polyps, nodules, cancer, among others diseases), neurological (dysphonia due to Parkinson's disease), and functional (hyper nasality due to cleft lip and palate).The results obtained with the stability and periodicity features indicate that these measurements are suitable to discriminate between healthy speakers and people with different kind of pathologies. This methodology uses different voice disorder dataset, get extracted with the features of each voice signal considering the duration and frequency, pad it, if necessary according to the consideration of the signal, train it and display the value in decimal format and then classify it through the classifier and test it lastly obtain the accurate percentage value of the classifier. For future work many other features extracting methods can be adapted and many more classification technique can be involved which will be useful for medical professionals for early detection of the disease and more dataset of English or other Indian religion languages can be involved.

REFERENCES

[1] A. Al-Nasheri et al., "Voice Pathology Detection and Classification Using Auto-Correlation and Entropy Features in Different Frequency Regions," in IEEE Access, vol. 6, pp. 6961-6974, 2018.
[2] M. Dahmani and M. Guerti, "Glottal signal parameters as features set for neurological voice disorders diagnosis using KNearest Neighbors (KNN)," 2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP), Algiers, 2018, pp. 1-5.
[3] T. B. Ijitona, J. J. Soraghan, A. Lowit, G. Di-Caterina and H. Yue, "Automatic detection of speech disorder in dysarthria using extended speech feature extraction and neural networks classification," IET 3rd International Conference on Intelligent Signal Processing (ISP 2017), London, 2017, pp. 1-6.
[4] J. R. Orozco-Arroyave et al., "Characterization Methods for the Detection of Multiple Voice Disorders: Neurological, Functional, and Laryngeal Diseases," in IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 6, pp. 1820-1828, Nov. 2015.
[5] L. Verde, G. De Pietro and G. Sannino, "Voice Disorder Identification by Using Machine Learning Techniques," in IEEE Access, vol. 6, pp. 16246-16255, 2018.
[6] S. Firdos and K. Umarani, "Disordered voice classification using SVM and feature selection using GA," 2016 Second International Conference on Cognitive Computing and Information Processing (CCIP), Mysore, 2016, pp. 1-6.
[7] K, Uma & Holi, Mallikarjun " A hybrid model for neurological disordered voice classification using time and frequency domain features" Artificial intelligence Research 2016, Vol. 5, No.1.
[8] E. B. Hook, "Congenital malformations worldwide: A report from the international clearinghouse for birth defect monitoring systems," American Journal of Human Genetics, vol. 51, no. 4, p. 919, 1992.

[9] D. Sell, P. Grunwell, S. Mildinhall, T. Murphy, T. Cornish, D. Bearn, W. Shaw, J. Murray, A. Williams, and J. Sandy, “Cleft lip and palate care in the United Kingdom–The Clinical Standards Advisory Group (CSAG) study. Part 3: Speech outcomes,” Cleft Palate-Craniofacial Journal, vol. 38, pp. 30–37, 2001.

[10] J. Hillenbrand and R. A. Houde, “Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech,” Journal of Speech and Hearing Research, vol. 39, no. 2, pp. 311–321, 1996.

[11] Crovato, Cesar & Schuck Jr, Adalberto “The Use of Wavelet Packet Transform and Artificial Neural Networks in Analysis and Classification of Dysphonic Voices” IEEE transactions on bio-medical engineering, vol. 54, pp. 1898-900, 2010.

[12] M.S Hossain, “Patient state recognition system for healthcare using speech and facial expressions,” Journal of medical systems, vol. 40, no. 12, pp. 272, 2016.