A TWO LEVEL ALGORITHM FOR DETECTION OF VOICE PATHOLOGY USING RANDOM FOREST WITH REDUCED FEATURE DIMENSION

Dr.G.L.Murthy1, R.Murali2, S.Vijaya Shilpa3, Ch.Durga Prasad4, D.Mounica5
1Professor, Department of ECE, Lakireddy Bali Reddy College of Engineering, Mylavaram
2,3,4,5 B.Tech Students, Department of ECE, Lakireddy Bali Reddy College of Engineering, Mylavaram
murthyfromtenali@gmail.com

Abstract
Voice pathology is the deviation from normal voice quality, pitch and loudness that are appropriate for given gender, age and culture. These voice disorders will have an adverse effect on vibration regularity and functionality of voice. Voice disorders are generally due to formation of extra tissue, inflammation and swelling of vocal cords, hormone disorder as well as excessive stress while speaking. Subjective evaluation is time consuming and is prone to error. The current work proposes a two level algorithm with an objective to isolate healthy controls from the pathological samples and then classifies samples based on their disorder type. The advantage lies in the fact that the features fed to the classifier are reduced there by which minimizing the computational cost.

Keywords: Dysphonia, Laryngitis, Mean of TEO, Absolute Mean Value, Random Forest

I. INTRODUCTION
Humans use voice as a primordial natural technique for communication. We used to rely on voice communication in both our personal and professional lives. The production of voice is the consequence of a complicated mechanism involving several respiratory organs. The sound produced by the vibrating of the vocal folds of the larynx is also known as the voice box, is referred to as voice. The throat and mouth modify this sound is known as the larynx-fundamental tone, to form speech. In recent years, a change in lifestyle has been linked to an increased incidence of pathological voice disorders. Around a quarter of the population works in jobs that are vocally demanding. For these people, either their occupations involve a lot of talking or their work settings cause them to talk over a lot of noise. Teachers, attorneys, auctioneers, aerobics instructors, singers, actors and manufacturing supervisors are among the professions with high vocal demands [1].

Voice diseases impact the vocal folds, causing irregular vibrations as a result of various causes that contribute to voice vibrations malfunctioning. Due to their partial closure, vocal fold diseases cause changes in the vibratory cycle of the vocal folds. Voice diseases also change the structure of the vocal tract and cause spectral abnormalities. The vibration of the vocal folds is affected by a variety of elements, including mucus on the tissue, stiffness, tension, muscles in the larynx and fold closure and opening.

The muscles inside the vocal folds endure sudden involuntary movements termed spasms, which interfere with the folds capacity to vibrate and create voice in spasmodic dysphonia. Anyone can be affected by spasmodic dysphonia. It's a rare condition that affects one to four people out of every 100,000 people. People between the ages of 30 and 50 are most likely to experience the initial signs of spasmodic dysphonia. It has a greater impact on women than on men.

Laryngitis is a condition in which the larynx swells and becomes inflamed. It might be acute or chronic, although most of the time it is only temporary and has no major implications.

Numerous voice problems, such as laryngitis, vocal cord paralysis, and dysphonia, have been studied extensively. Linear Prediction Coefficients (LPC), Mel Frequency Cepstral Coefficients, Pitch, and Degree of Vocalization were among the criteria utilized to detect abnormalities. Following the extraction of these features, classification methods such as Support Vector Machine (SVM), K-nearest neighbor algorithm, and Random Forest Algorithms
are used to classify the voice samples. However, the above methodologies are generally reported to have limitations such as decreased accuracy with increasing data and a restriction to a particular voice problem.

The current effort aims to develop a high breed algorithm to recognize multiple vocal problems. Teager Energy Operator (TEO) selects notable aspects from the core properties of speech signal meant for creating it. It makes use of standard Saarbrucken Voice Database (SVD) that is a collection of voice recordings of various illnesses, both functional and organic, that is freely available online. A collection of over 2000 people's voice recordings is maintained in the database. It comprises of recordings of the vowels [i, a, u] generated at three different pitches they are normal, high, and low. The files with sustained vowels are between 1 and 3 seconds long. All recordings are captured at 50 kHz and have a 16-bit resolution. There are 71 different diseases in total, including functional and organic. After removing those where some of the recordings were lost or damaged, a total of 1970 sessions are maintained. Pathological speakers account for 1320 (609 males and 711 females) sessions, while normal speakers account for 650 (400 males and 250 females).

II. EXISTING WORK

In [2], Ugo Cesari et al., discussed about a system for estimating the fundamental frequency, jitter, shimmer and harmonic to noise ratio (HNR). These parameters play a vital role in the process of diagnosing the voice disorders. Android SDK is used as the platform for acquiring the voice samples and evaluation of characterizing parameters thereby. However, the analysis is constrained to the parameters covered in the questionnaire that necessitated further examination. In [3], Joao Paulo Teixeira et al., considered set of features for statistically analyzing the Laryngitic patients and healthy controls. The work is fundamentally aimed at developing a classifier based on either SVM or ANN with improved accuracy. It was concluded that the gender plays no prominent rule in the process of discrimination. In [4] Laura Verde et al., highlighted the significance of mobile devices for health monitoring. Numerous machine learning algorithms have been compared with reference to dysphonia. Either SVM or decision tree algorithms yielded best results based on the selected features. D Hemmerling [5] used the auto associative neural networks for identification of voice related pathologies. A feature set of 35 parameters has been used for numerous voice disorders and the results were quite promising. Gulam Mokhmed et al., [6], have made use of parameters taken from the vocal tract area in developing a system for detecting voice pathology. It was mentioned that irregular patterns will be exhibited as vocal tract is physically connected to vocal cords. Hamzeh Ghasemzadeh et al., [7] mentioned that speech production mechanism follows non linearity. In such a scenario, instead of relying on linear parameters, the non linear features vice mutual information, false neighbor fraction and Lyapunov spectrum are considered for pathology detection. The computational complexity is reduced by making use of feature reduction algorithm Linear discriminant analysis (LDA).

In [8] S.Jothilakshmi has proposed an algorithm for classifying a broad range of voice disorders by making use of GMM and HMM classifiers. The discrimination of pathological disorders is carried out by eliminating the normal controls from the classification process. Ahmed al nasheri et al.,[9] developed an algorithm for for pathology detection by analyzing the frequency bands. Given voice sample is divided into number of frames and the features chosen are correlations functions.FIR filters with numerouscentre frequencies are used to analyze the voice signal. However, there existed variability in the accuracy from one data base to another. In [10], Mazin Abed Mohammed et al., improved the accuracy of pathology detection by using CNN model with prior training. Reducing the duration of training with improved precision is achieved by the incorporation of transfer learning. ResNet34 model is used, as the residual network is one of the most prominent CNNs. Pavol Haror et al., [11] included recurrent long term short term memory for detecting voice pathologies. It was mentioned that high end multi layer networks can be developed due to the availability of advanced deep learning models.

The current work proposes a two level algorithm for identifying the pathology in a given voice signal. Initially the healthy samples are separated from the pathological samples and then discrimination is made between nature of pathology. The following flow diagram will visualize the various process steps involved in the disorder detection process.

III. PROPOSED WORK

The non stationary of the real time signals made it mandatory to process them in blocks or frames to emphasize more on the local features than opting for global. Indeed, the behavior is not same throughout the sample there by which yielding in accurate results. Fig. 1 visualizes various steps involved in the current work. In preprocessing, the speech signal is divided into number of frames by means of windowing. The consequence of
transforming a large sequence of speech samples into constituent frames should not give more side lobes thereby resulting loss of energy. Among numerous windows available, Hamming window is used in the process of localizing the parameter variation. This resulted in number of frames comprising of 260 samples with overlapping. Fig.2 presents a portion of healthy sample along with samples of the selected pathologies for classification dysphonia and laryngitis.

3.1 Mean of Teager Energy Operator (MTEO)

To reduce the number of parameters meant for identification of voice disorder, the proposed work will isolate the pathological voice samples from non pathological. The pathologies that were considered in the current work are dysphonia and laryngitis. This discrimination is carried out by applying the TEO to each frame and thereby estimating the mean of Energy for each frame.

The Teager Energy Operator (TEO) primarily displays the frequency and instantaneous changes in the signal amplitude, which is extremely sensitive to minor variations. TEO was first developed for modelling nonlinear speech signals, but it has since become widely used in audio signal processing.
The Teager energy operator can be used to track many information sets in voice signals, for example. In the voice signal, this operator may track vowels and formants. It is also used to figure out the formants central frequency and bandwidth. The Teager energy operator has recently been used in the implementation of a vocal pathology detection method.

\[
\varnothing[x(n)] = x^2(n) - x(n+1)x(n-1)
\]

Where,

\[
\varnothing[x(n)] = \text{Energy operator}
\]

\[x(n) = \text{Discrete time signal}
\]

\[n = \text{Sample size}\]

It is observed that Mean of TEO (MTEO) for pathological samples is deviating from the healthy controls. 126 samples from SVD database are taken out of which 42 are healthy controls and 42 samples with dysphonia and 42 with laryngitis. Fig 3 exhibits the TEO profile of all three cases individually and merged in one also. The estimated Mean TEO profile of all the samples is fed to random forest classification. The proposed algorithm has incorporated 100 trees in the process of classification. The classification accuracy was 92.3% with the test sample size as 20% of the total number of samples. From the confusion matrix, that is meant for describing the performance of classification algorithm, it was found that only two samples are misclassified as healthy even though not so.

![TEO profile of different cases](image)

**Feature extraction:**

After narrowing down the sample space by separating the pathological and healthy controls, the next task is to distinguish between the nature of pathology. Two voice disorders, effort has been put to classify the non pathological voice samples into disphonia and laryngitis. Two prominent voice features Mel-Frequency Cepstral

[www.turkjphysiotherrehabil.org](http://www.turkjphysiotherrehabil.org)
Coefficients (MFCC) and linear predictive coding (LPC) with reduced feature dimension are selected as the prominent features.

**Mel-Frequency Cepstral Coefficients (MFCC)**

In voice disability detection algorithms, the MFCCs have been frequently used. MFCCs have the benefit over other voice features in that they can fully characterize the geometry of the vocal tract. Once the vocal tract has been adequately defined, an accurate depiction of the phenomena produced by the vocal tract may be estimated. The MFCCs accurately capture the envelope of a short-time power spectrum, which represents the geometry of the vocal tract. By converting the conventional frequency to Mel Scale, MFCC takes into account human perception for sensitivity at appropriate frequencies, and are thus ideally suited to speech recognition tasks, as they are well suited to understanding humans and the frequency at which humans speak/utter.

\[
x(n, w_k) = \sum_{m=-\infty}^{\infty} x[m]W[n - m]e^{-j\omega_km}
\]

Where,

\[
w_k = 2\pi k \div N
\]

\[
x(n, w_k) = \text{weighted by a series of filter frequency responses}
\]

\[
N = \text{Discrete Fourier transform length}
\]

**3.2.2. Linear Predictive coding**

LPC was created primarily to compress digital signals for faster transmission and storage. However, LPC has grown in prominence as a formant estimator and has become one of the most powerful speech analysis approaches. The LPC approach uses a linear all-pole infinite impulse response (IIR) filter to represent the vocal tract.

Linear predictive coding (LPC) is a method for compressing the spectral envelope of a digital voice signal using information from a linear predictive model. It is widely utilized in audio signal processing and speech processing. In voice coding and synthesis, LPC is the most extensively used approach.

\[
V(z) = \frac{A}{1 + \sum_{k=1}^{p} a_p(k)z^{-k}}
\]

Where,

\[
p = \text{Number of poles}
\]

\[
A = \text{Gain of the Filter}
\]

\[
a_p(k) = \text{Filter Coefficients.}
\]

Mel frequency cepstral coefficients that reflect the power spectrum for a shorter duration are estimated frame basis. For each frame, 12 coefficients are estimated that results in a large dimension space with 198 frames and each with 12 coefficients. To reduce the cost of computation, selection of prominent coefficient and Fig. 4 depicts the variability among MFCC coefficient (MFCC9) for all the frames. Fig. 5 evidences the reason for selecting a specific frame for the purpose of classification. It can be seen that the amplitude variations in dysphonia is comparatively smaller than that for laryngitis. The absolute mean value of MFCC9 for all the frames is estimated and used as one feature for the classification.

Second order predictor is used in the current algorithm and the resulting coefficients are analyzed for variability. The n+1 coefficient exhibited much variance when compared to other coefficients in regard to dysphonia and laryngitis. Table 1 depicts the 2nd order LPC derived for specific sample.

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathology</td>
</tr>
</tbody>
</table>

www.turkphysiotherrehabil.org
Thus $3^{rd}$ LP coefficient is chosen as another feature for the purpose of classifying the pathology. These selected features are applied as input to random forest classifier. It is an enhancement of decision tree algorithm that constructs a population of decision trees during training time for classification. The prominence of random forest lies in the fact that it tries to minimize the variance by choosing samples from different parts of training sets.

### IV. CONCLUSION

The objective of the feature dimension for classification is achieved in the current work by a two level mechanism. Isolating the healthy control from pathological samples at the fundamental level is done by using TEO profile. The classification accuracy stood at 92.3% when random forest algorithm with 100 decision trees is applied to 126 samples. From the confusion matrix it is observed that only two pathological samples are misclassified as healthy meaning that number of false negatives is two. From the confusion matrix, the precision stood at 100 % while the recall rate was 77% for the 26 test samples chosen. Once the healthy controls are separated, MFCC and LPC are given as input to the classifier, accuracy of 94.11% is derived with only one misclassification. The feature dimension size is drastically reduced by selecting the prominent frame as well as coefficient.

<table>
<thead>
<tr>
<th></th>
<th>Dysphonia</th>
<th>Laryngitis</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP coefficient</td>
<td>-1.5772</td>
<td>-1.0380</td>
</tr>
<tr>
<td></td>
<td>0.5891</td>
<td>0.0459</td>
</tr>
</tbody>
</table>
V. REFERENCES


