Detection And Classification Of Speech Pathology Using Deep Learning

Dr. S.Santhana Megala, Dr. R. Padmapriya, Dr. B.Jayanthi, Dr. M.Suganya,

Abstract: Speech or Voice Pathology investigation performs a significant role in the recent record of Health Industry. The need for analysis is that the detection and classifications of tones of pathological voices, till now, which is believed as a tough task within the sector of speech analysis. Sometimes Patients are probably in tough state to identify a modification in voice parameters, like hoarseness; however the voice pathologies might result from a large spectral fluctuate of causes, like respiratory disease to a cruel tumor. Medical practitioners like otolaryngologists were discovering different kinds of speech pathologies from the patient's speech from mouth i.e. oral communication. Unluckily, this classification rate of Speech pathology by the physician consultants is just concerning 60-70%. Thus tone of voice or speech pathologies is found by the analysis techniques like laryngostroboscopy or small laryngoscopy, which distress the individual to a decent scope, in addition to that it is expensive. It is not possible to detect the speech pathology at the initial stage by manual diagnosis. The primary objective of this paper is to propose automatic diagnostic tools to assist the voice or speech diagnosis. This speech pathology identification system works based on the support of the medical practitioner, which helps in identifying the pathology even in the beginning stage. In this paper, the speech or voice signal is examined by the acoustic variables like Noise removal, Windowing, Smoothing, Mel consistency and Jitter. Finally the classification of voice pathology is done based on the Neural Networks and Deep Learning. The experimental results were also discussed in detailed manner. From the experimental results it is clear that the Speech pathology recognition system successfully classified and labeled the normal voice and the pathological voice.

Index Terms: Voice Pathology, Classification, Neural Network, Mel-Frequency Cepstrum, Deep Learning.

1 INTRODUCTION

Within the recent history of medical diagnosis, a substantial interest was shown in field of voice and speech analysis, which basically an analysis of the patient's voices, who have a draw back inside the vocal cords. In analyzing the patient's voice, a voice therapist desires some intensive trail and error analysis method to find the problem. It is terribly exhausting in every state of affairs, as a result of voice problems typically result from the vocal folds or organ muscle, which controls the voice created by the human. Such fold analysis and to look at their movements is physically a hard one, therefore there is a need for a tool or instrumentation to check the tone of speech or voice pathology. This need paves a way to the researchers to look out some technology to help the voice therapist find the voice pathology. Such technological support is carried out to support the whole process in finding and analyzing the foremost frequent vocal disorders, their symptoms, and root causes and their side effects. This supported researcher to perform an analysis to create a voice pathology identification system, which will be a moderate, non- intrusive, and stable processed system used for the recognition of pathologies inside the human's vocal cords.

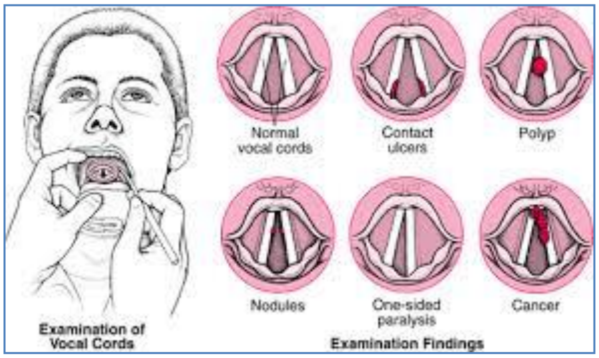


Fig. 1: *Examination of Vocal Cords to find the Pathology*

Acoustic attributes plays as a main parameter to discriminate the tone of speech or voice impulses as normal or pathology. Acoustic attributes were used to evaluate the options or physical characteristics of voice communication that transmitted at every illustration of time. By applying the Predication techniques on the attributes, extracted from the Acoustic Feature. Such processes analyze and verify the tone of voice or speech signal whether it is created from the vocal fold, which is suffering by some pathology or not. This paper explores a proficient system which detects and classifies the Voice Pathology. The dataset is developed by documenting the human tone of voice throughout a sound and noises free surroundings. The voice communication indication is then analyzed to prolong or deduct the acoustic parameters like the Noise removal, Windowing, Smoothing, Mel Frequency and Jitter. Following, to the feature Extraction process, the discrimination of types of voice pathology based on the features extracted from the acoustic variables is classified using the popular techniques such as Neural Networks and Deep Learning. Finally a well tuned processed Voice Pathology Detector was designed, which can facilitate the expert to investigate and realize the patient's voice. This paper

- *Dr. S. Santhana Megala, Assistant Professor in Department of BCA, santhanamegala@rvsgroup.com*
- *Dr. R. Padmapriya, Associate Professor & HoD – BCA, padmapriya@rvsgroup.com*
- *Dr. B. Jayanthi, Associate Professor & HoD – CS, Jayanthi@rvsgroup.com*
- *Dr. M.Suganya, Associate Professor & HoD – IT, suganya@rvsgroup.com*
  School of Computer Studies, Rathnavel Subramaniam College of Arts and Science (Autonomous,) Sulur,, Coimbatore, Tamil Nadu, India.

showcases the previous research works done in the literature review section 2. The proposed methodology of the research work is described in section 3. The implementation framework and the experimental results were mentioned in section 4. Finally, the ultimate outcome is described in the section 5.

2 LITERATURE REVIEW

Detection of voice or speech pathology and its treatment process were adopted based on the medical practices which are learned through some investigations done by the research experts and by gaining knowledge from the speech-language pathologists [1]. The speech pathology were analyzed in [2], the constant strategies were found in distinctive speech organ which removes transmission in speech impulses, but the repetition time rate , magnitude and modulation triggers were predicated within the Non constant waysThe [3] foremost widespread Acoustic feature extraction approach, such as MFCC, Mel-Frequency Cepstrum Coefficients were utilized as basic process to classify the normal and pathology voice or speech pattern based on the GMM Classifier. The [4] insight voice datasets accustomed to get the pathology present in it were compressed and held up in MP3 format and thus the Classification Methodologies familiarized to classify the normal and pathology voice or speech, which were examined, analyzed and compared among the list of techniques such as GMM and SVM. The [5] classification techniques such as PCA, Primary Component analysis and SVM, Support Vector Machine were used to classify the normal and pathological speech predicated based on the 27 important features extracted from the real time human voice or from speech dataset. In [6] alternating process was initiated by selecting the Hereditary rule to spice up the features framed from the speech signals. Further the SVM Kernel classification techniques were used to classify the speech pathology.

3 PROPOSED RESEARCH METHODOLOGY

The primary objective of the proposed work is to provide an automatic voice analysis system, which could be used to find if the patients voice consists of pathology voice disease or not. Sometimes patients visit the hospital and look forward for the appointment of the medical experts to look at the tone of voice and sit down with them for the evaluation. If a person contains some voice pathology disease, he may wait to get medication to cure the disease. Suppose, if a normal person is having a doubt, that he is having a pathology in his or her voice, then it might be a difficult one and a frustrating one to wait for a long time. Thus the need for the pathology detection technique arises. This research work proposed a design to detect and analyze the voice pathology in a simple way. The proposed voice or speech pathology identification system is framed based on three phases such as Pre-processing, Feature Extraction and Classification. At first the data set were collected from MEEI Dataset and the noises present in the voice were removed using the pre-processing techniques such as Noise removal, and Windowing. Then the features that contained in the voice signal were find out to classify the voice signal such features were, Smoothing, Mel Frequency and Jitter. Finally, the discrimination of types of voice pathology based on the features extracted from the acoustic variables is classified using the popular techniques such as Neural Networks and Deep Learning. Finally a well tuned processed Voice Pathology Detector was designed, which can facilitate the expert to investigate and realize the patient's voice. The phase division for the proposed research work is described in figure 1.

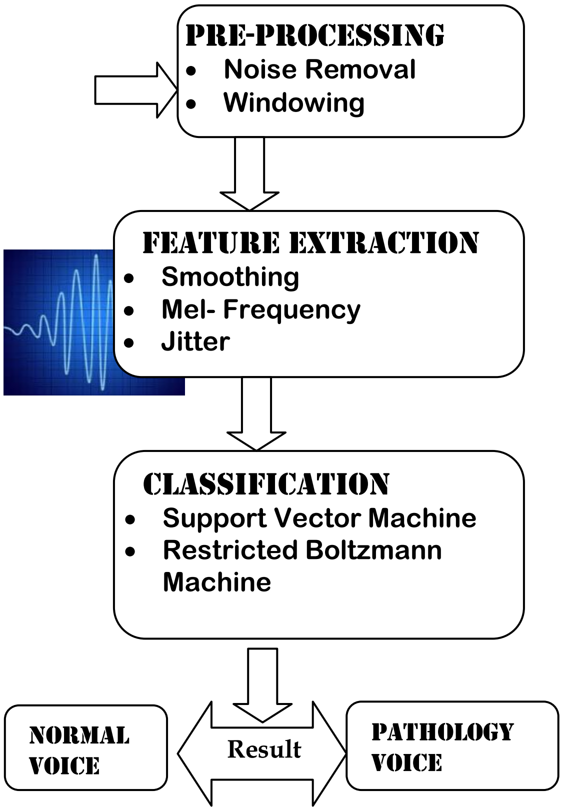


***Fig. 2**: Research Plan for the proposed work*

3.1 Pre-Processing

Pre-processing phase is considered as the first phase in the detection of voice pathology process. It is used to recognize the signal to differentiate the voiced or unvoiced signal and to create feature vectors from the signals. Preprocessing techniques adjusts or modifies the voice signal, i.e. x(n), in an acceptable format for further processing.

Noise Removal:

The voice signals in the transmission are often go along with a lot of unnecessary disturbing noise. Typically, sound absorbing cotton or directional microphones were covered on the hardware to hold back the background noise that raised by the unwanted signals, such as sound of wind and movement of objects. The major issue in noise removal, when it comes to speech signal processing. The process is to check the speech signal, x(n). If the speech signal is corrupted by some background or ambient noise, such as, d(n), which is mentioned as additive disturbance. This can be removed by using the equation 1.

$$x(n) = s(n) + d(n) \quad (1)$$

From the above equation, s(n) is the clean speech signal. Likewise there are different noise reduction methods that can be adopted to perform the task on a noisy speech signal.

Windowing: In this stage, the speech or voice signal has been framed into segments. Each frame in this signal is multiplied with a window function i.e. w(n) with length denoted as N, where N is the length of the voice signal frame. In signal processing, Windowing is a process of multiplying a voice

signal segment in wave form, by a time window for a specified shape, to strain pre-defined distinctiveness of the signal.
Hamming Window:

$$\boldsymbol{w(n) = 0.54 - 0.46\, Cos(2n\pi/(M-1))}, \quad \boldsymbol{0 \leq n \leq M-1} \quad (2)$$

To minimize the discontinuity of voice signal at the beginning and at the end of each frame, the voice signal should be lessened to zero or close to zero, and therefore it minimizes the mismatches. To acquire good frequency in signal resolution, a long window is advisable but the importance of some short transmission makes a short window desirable and effective. A normal compromise in the quality of the signal, that is always available to patch up if the frame length of the signal is about 20 or 30 ms, and with a frame spacing of 5 to 10 ms.

3.2 Feature Extraction

Feature Extraction techniques may act upon a significant role in selecting the features that present in the voice signal, which is well appropriate for the classification process of voice pathologies. For a Decision Making System, Feature Selection process is an important one. Let us see the techniques discussed in this section. MFCC i.e., Mel Frequency Cepstral Coefficients, is one among the popular feature extraction techniques that available for the signal processing field. In speech analysis, Mel Rate of Occurrence Cepstrum is a representation of a brief -term power spectral range of a voice transmission, which made and predicated on the linear circular function amendment of the log power varies over a nonlinear Mel-scale rate of prevalence. Mel Rate of recurrence Cepstral coefficients became a member of signed up with as remove a merged group to constitute an MFC. The Mel Frequency Cepstrum uses spaced frequency bands on the Mel Scale equally, that calculate approximately the human auditive system's response. It additionally evaluates the quality of Cepstrum that uses linearly-spaced frequency bands in it. Jitter is a techniques used to check the dimension reduction on the vocal stability. Actually, the speeds of repetation of the speech transmissions can be improved from one frame to another frame. By locating the random data variability, vocal interference are often calculated, that is responsible for gruff or rough voice signal. The interference rate of repeated variability for a unique tone of voice is considerable for evaluation. Therefore interference could be a trusted approach for speech pathology analysis.

3.3 Classification

3.3.1 Levenberg-Marquardt Algorithm

The Levenberg-Marquardt algorithm, is designed to work distinctively also known as the damped least-squares method, has been designed to work particularly with loss functions which take the form of a sum of squared errors. It works without computing the exact Hessian matrix value. As an alternative way, it works with the gradient vector and the Jacobian matrix. The loss function can be expressed as a sum of squared errors of the form of,

$$f = \sum_{i=1}^{m} \boldsymbol{e_i^2} \quad (3)$$

The Levenberg Marquardt algorithm also approaches the Newton method which typically accelerates the convergence to the bare minimum one.

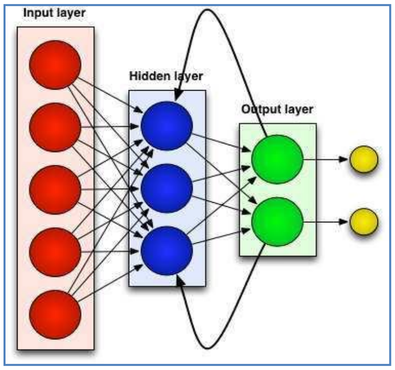


***Fig. 3**: Layer overview of Levenberg-Marquardt algorithm*

3.3.2 Restricted Boltzmann Machines

Boltzmann Machines is a technique used in the form of log linear Markov Random Field (MRF). In which the energy function is linear in its parameters [4]. Hidden nodes were introduced to make them powerful enough to represent complicated distribution. By introducing more hidden variables, one can increase the modeling capacity of the Boltzmann Machine. Further, Restricted Boltzmann Machines are Boltzmann Machines, which does not contain visible-visible and hidden-hidden connections [4]; therefore the name came as 'restricted'. A graphical representation of an Restricted Boltzmann Machines is shown below.

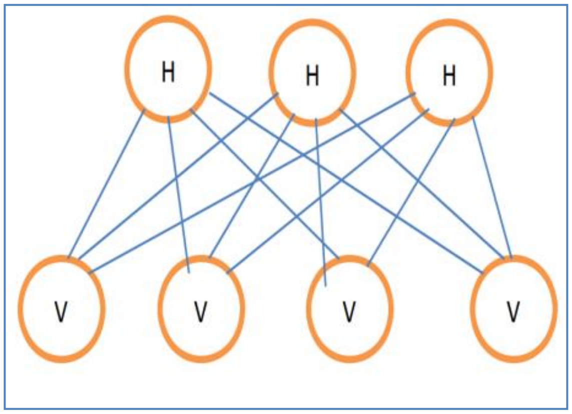


***Fig. 4**: A graphical representation of Restricted Boltzmann Machines.*

The Energy function E (v, h) is given by equation (4):

$$\mathrm{E(v,h) = -\ b'v - \ c'h - \ h'Wv} \quad (4)$$

where W represents the weights connecting hidden and visible units and b, c are the offsets of the visible and hidden layers respectively. In context to free energy, equation (5) is given by:

$$\mathrm{F\ (v) =-\ b'v - \sum i \log \sum e^{h} i^{c}\ i^{+w} i^{v}} \quad (5)$$

Sampling of RBM can be done by running a Markov chain to convergence, which uses Gibbs sampling as the transition operator.

4 EXPERIMENTAL RESULTS

The detection and classification of voice pathology was designed and implemented in Matlab version 2016a.

Table 1: *MEEI Dataset Description*

			Male	Female	Total
Pathology	Physiological disorder	Vocal Fold nodules	1	18	19
		Vocal Fold edema	9	31	40
	Neuromuscular disorder	Vocal Fold Unilateral paralysis	29	30	59
Normal	Healthy Voice		14	22	36
Total			53	101	154

The MEEI database [10] contains 53 healthy subjects and 724 subjects with voice pathologies. For these two sets of subjects, the sustained vowel /a/ and a continuous reading speech excerpt, "rainbow passage" are available. From the database only 477 subjects have the pathology information.

This study uses subjects with nodules, edema and unilateral vocal fold paralysis and healthy subjects.

The MEEI database [10] contains 53 healthy subjects and 724 subjects with voice pathologies. For these two sets of subjects, the sustained vowel /a/ and a continuous reading speech excerpt, "rainbow passage" are available. From the database only 477 subjects have the pathology information.

This study uses subjects with nodules, edema and unilateral vocal fold paralysis and healthy subjects.

This paper used the MEEI database [10] for experimental purpose which contains 53 subjects with normal condition and 724 subjects with voice pathologies. For this test set two sets of subjects were trained based on the sustained vowel /a/ and a reading speech of words, "rainbow passage". From the above database, here a sample of 118 subjects having the pathology in their voice, specifically with nodules, edema and unilateral vocal fold paralysis along with 36 normal subjects were used for testing.

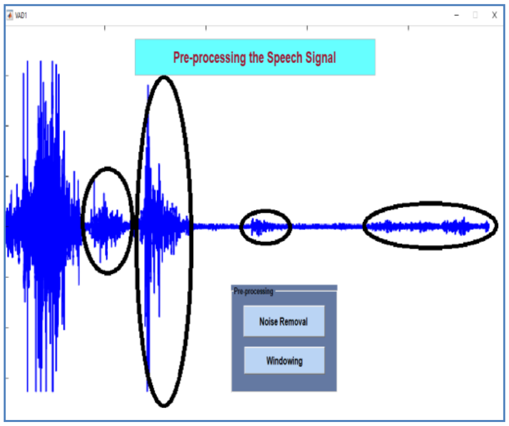


Fig. 5: *Pre-processing the Speech Signal*

In this Pre-processing phase, the sample speech .wav file is pre-processed by implementing the noise removal method and Windowing techniques.

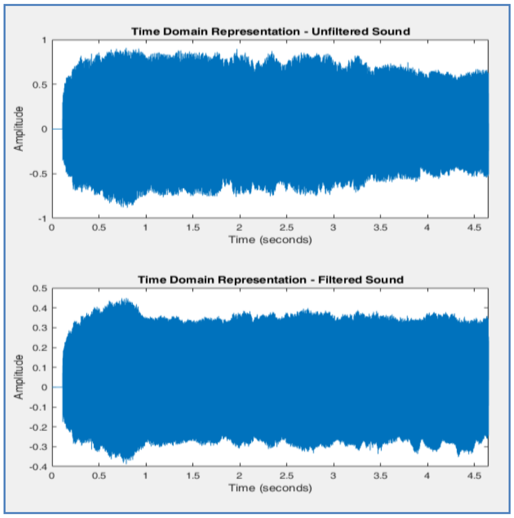


Fig. 6: *Noise Removal in Time domain Representation*

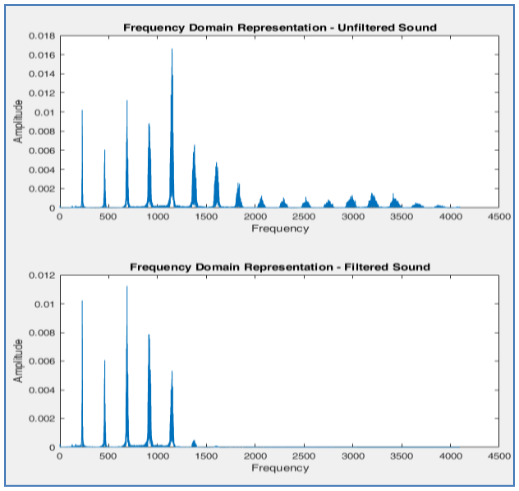
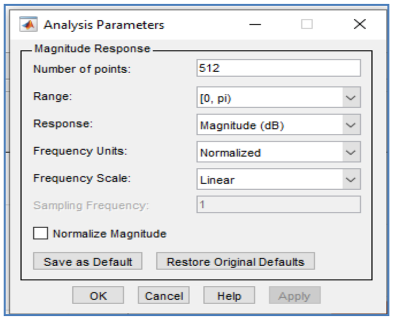


Fig. 7: *Noise Removal in Frequency domain Representation*

The Noise removal is done in two domain such as Time domain and in frequecy domain. The time domain representation of noise removal is depicted in figure 6. The

Frequency domain representation of noise removal is depicted in figure 7.



Analysis Parameters

Magnitude Response

Number of points: 512

Range: [0, pi)

Response: Magnitude (dB)

Frequency Units: Normalized

Frequency Scale: Linear

Sampling Frequency: 1

☐ Normalize Magnitude

Fig. 8: *Analysis Parameters for Windowing process*

In windowing technique, the speech or voice signal has been framed into segments. It multiplies a voice signal segment in wave form, by a time window for a specified shape, to strain pre-defined distinctiveness of the signal. Figure 8 shows the parameter values needed for the analysis of windowing process. Figure 9 shows the Normalized Frequency domain in Magnitude. Figure 10 shows the Window samples in time domain.

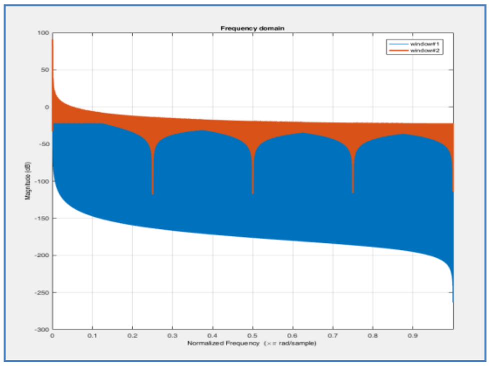


Fig. 9: *Normalized Frequency domain in Windowing*

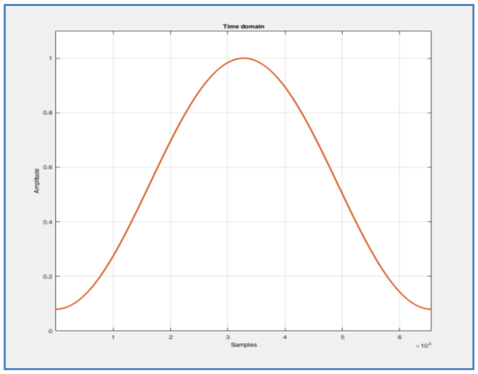


Figure 10: *Window sampling in Time domain*

The phase II contains the feature extraction techniques to uniquely identify the features that present in the speech or voice signal. Mel Frequency Cepstrum and the Mel Frequency Cepstrum Coefficients were used in this phase, which equalizes the input speech or voice frequency measure. It depicts the ability of the spectral envelope of single frame. In this paper, 12 Mel Frequency Cepstrum Coefficients were generated from the smoothened signal using filtering techniques. Figure 11 shows the framework designed for the Feature Extraction phase.

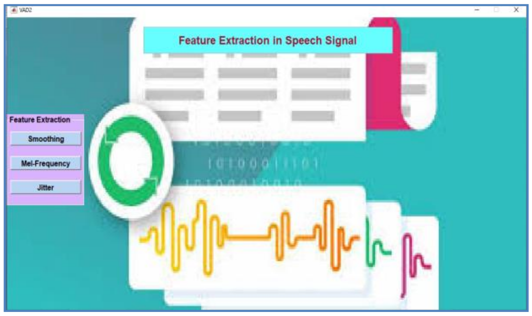


Fig. 11: *Framework for Feature Extraction*

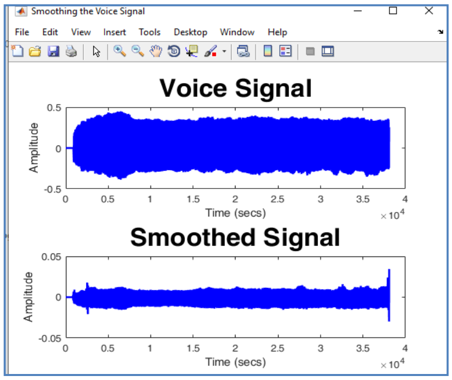


Figure 12: *Smoothing the Voice Signal*

Figure 12 depicts the normal voice signal and smoothened voice signal. Figure 13 & 14 represents the Mel – Frequency Cepstrum coefficients view.

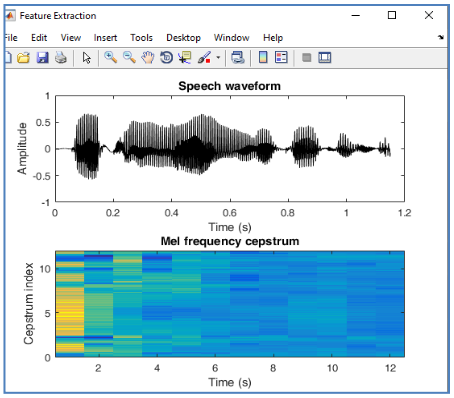


Figure 13: *Normal speech waveform vs Mel Frequency Cepstrum*

The normal input speech waveform is displayed in the first part after removing the noise that present in it and the Mel frequency cepstrum is displayed in the second part.

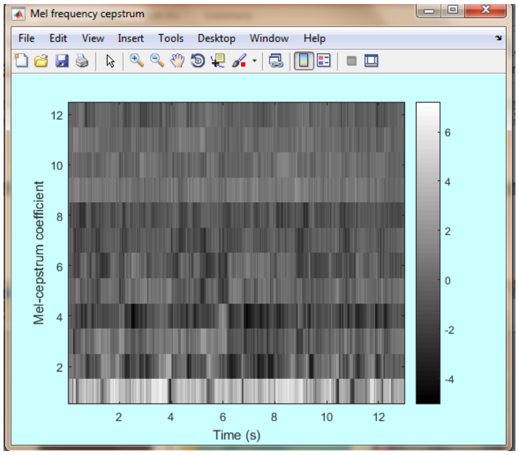


Figure 14: *Twelve Mel-Frequency Cepstrum*

Neural Network and deep learning techniques were implemented to classify the speech or voice signals to check whether the voice signal is the normal or pathology. A popular Neural network algorithm such as, Levenberg Marquardt Algorithm is used for classification of voice signals. On the other hand, Restricted Boltzmann Machine algorithm is used to implement the Deep Learning classification of voice signals. The classified results were segmented as normal and pathology, if pathology it will display the diseases affected by the patient based on the measurements. The disease taken from the database such as vocal Fold nodules, vocal fold edema or vocal fold unilateral paralysis for experimental study. Figure 15 & 16 represents the Classification of voice signal using Restricted Boltzmann Machine and its sample Normal voice result was discussed.

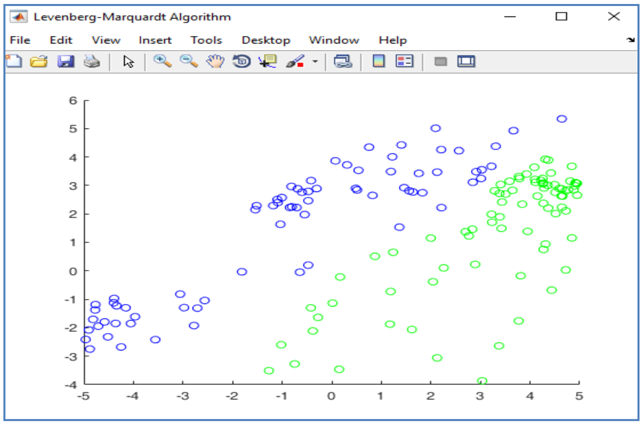
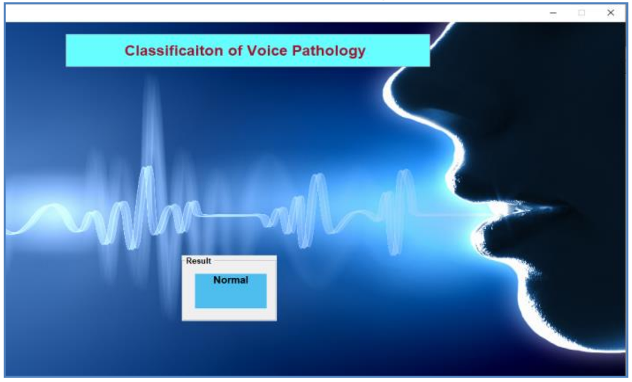


Fig. 14: *Classification of Normal Voice using Restricted Boltzmann Machine algorithm*



Classificaiton of Voice Pathology

Result

Normal

Fig. 15: *Sample result for a Normal voice*

Figure 16 & 17 represents the Classification of voice signal using Levenberg Marquardt algorithm and its sample Pathology voice result was discussed.

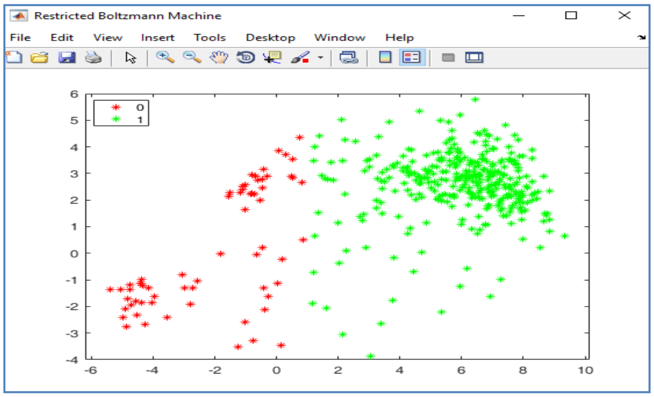
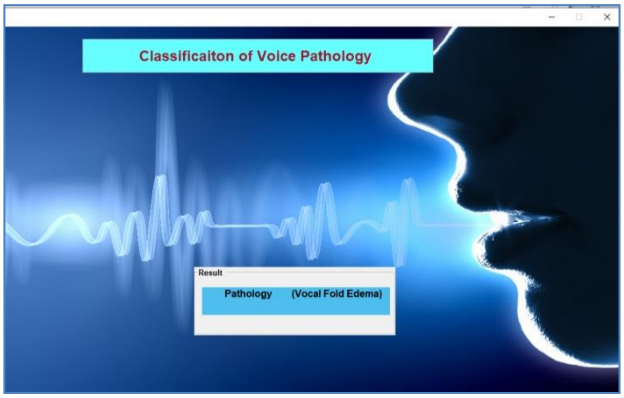


Fig. 16: *Classification of Pathology Voice using Levenberg Marquardt Algorithm*



Classificaiton of Voice Pathology

Result

Pathology (Vocal Fold Edema)

Figure 17: *Sample result for a Pathology voice*

Table 2: *Classification of voice signals using Levenberg-Marquardt and Restricted Boltzmann Machine based on Accuracy*

Algorithms	Accuracy
Levenberg Marquardt	92%
Restricted Boltzmann Machine	98%

Even though both techniques give actual results, there is a need to find the best methodology among them. Thus Accuracy measure is used to analyze both algorithms. Hence, Restricted Boltzmann Machine algorithm gives the accuracy of 98% which is better while comparing to Levenberg-Marquardt Algorithm which gives 92% Accuracy in predicting the Voice pathology.

5 CONCLUSION

In this paper, a Voice or Speech Pathology Identification System was designed and implemented using neural networks and Deep learning concept. Some sample .wav files from MEEI Dataset were taken for the experiments and results analysis. The Pre-processing techniques like noise removal, and windowing were done. The Feature Extraction Techniques such as Smoothing, Mel-frequency Cepstrum and Jitter were used to extract the useful features which will help in classifying the signals. Finally Classification techniques adopted here were Levenberg-Marquardt Algorithm and Restricted

Boltzmann Machine algorithm. Hence by using the hidden layer concept, the classification of voice analysis was done in an efficient manner. By analyzing the experimental results, it shows that the proposed work accurately processed and classified the patient's voice samples into normal and pathology for both the classifiers. In addition to that, Restricted Boltzmann Machine algorithm gives better accuracy while comparing to Levenberg-Marquardt Algorithm which has only one hidden layer, when compared to the Deep learning techniques. Thus further investigation in this research work was planned to proceed along with Deep Learning Techniques.

REFERENCES

[1] Ali, Alsulaiman, Muhammad, Elamvazuthi, Mesallam, "Vocal fold disorder detection based on continuous speech by using MFCC and GMM". 7th IEEE GCC Conference and Exhibition (GCC), Pp: 292 - 297, 2013.

[2] Alsulaiman, "Voice pathology assessment systems for dysphonic patients: Detection, classification, and speech recognition". IETE Journal of Research, Vol: 60, Is: 2, Pp: 156 - 167, 2014.

[3] Chang, and Lin, "LIBSVM: A library for support vector machines". ACM Transactions on Intelligent Systems and Technology, Vol: 2, Is: 3, Pp: 1 - 27, 2011.

[4] Geoffrey Hinton, Ruslan Salakhutdinov, "Deep Boltzmann machine", International Conference on Artificial Intelligence and Statistics (AISTATS - 2009), Vol:5, 2009.

[5] Huang, Zhou, Ding, and Zhang, ''Extreme learning machine for regression and multiclass classification,'' IEEE Transaction on Systems, Management and Cybernetics, Vol: 42, Is: 2, Pp: 513 – 529, 2012.

[6] Klára, Viktor, Krisztina, "Voice disorder detection on the basis of continuous speech", 5th European Conference of the International Federation for Medical and Biological Engineering, Vol:37, Pp: 86 - 89, 2012.

[7] Maria, Sever, and Carlos, ''Cloud computing for big data from biomedical sensors monitoring, storage and analyze,'' Conference on Grid, Cloud High Performance in Computer Science, Pp: 1 – 4, 2015.

[8] Markaki, and Stylianou, "Voice pathology detection and discrimination based on modulation spectral features". IEEE Transactions on Audio, Speech, and Language Processing, Vol: 19, Is: 7, Pp: 1938 - 1948, 2011.

[9] Mohammed, Far, and Naugler, ''Applications of the map reduce programming framework to clinical big data analysis: Current landscape and future trends,'' Bio Data Mining, Vol: 7, Is: 22, Pp: 1 – 23, 2014.

[10] Muhammad, ''Voice pathology detection using interlaced derivative pattern on glottal source excitation,'' Biomedical Signal Processing and Control, Vol: 31, Pp: 156 – 164, 2017.

[11] Muhammad, Melhem, "Pathological voice detection and binary classification using MPEG-7 audio features", Biomedical Signal Processing and Control, Vol: 11, Pp: 1 - 9, 2014

[12] Nepal, Ranjan, and Choo, ''Trustworthy processing of healthcare big data in hybrid clouds,'' IEEE Cloud Computing, Vol: 2, Is: 2, Pp: 78 – 84, 2015.

[13] Salhi, Mourad, and Cherif, "Voice Disorders Identification using Multilayer Neural Network," International Arabian Journal of Information Technology, Vol: 7, PP:177–185, 2010.

[14] Sonu, and Sharma "Detection of Disease Using Analysis of Voice Parameters'' International Journal of Computing Science and Communication Technologies, Vol:4, Is: 2, 2012.

[15] Tanaka and Okutomi, "A Novel Inference of a Restricted Boltzmann Machine," 22nd International Conference on Pattern Recognition (ICPR - 2014), Pp: 1526 - 1531, 2014.