ISSN No:-2456-2165

Application of Speech Processing for Pathological Voice Detection and Analysis

Smitha S M, Prema K N, Roopa B S
Asst. Professors Dept.of ECE, JNNCE, Shivamogga
Karnataka, India

Abstract:- In recent years, speech pathology analysis has played a big role in the announcements of doctors. so as to speak with an individual and categorical our thoughts and feelings, the speaker must act with the auditor by manufacturing speech input, that any is taken by the auditor and work is allotted consequently. The method by that thoughts area unit translated into speech involves the utilization of articulators for manufacturing numerous speech sounds. Speech defect may be a upset that's caused thanks to lack of management of assorted organs that area unit employed in the assembly of speech. The popularity and allocations of tone of pathological voices area unit believed as a difficult add the sector of speech analysis still currently. Which creates the requirement to search out some technology to assist the voice expert in police investigation the voice pathology? Such technological support may be done supported the entire understanding or learning of the foremost frequent vocal disorders, their symptoms, real cause and malady aspect results. Acoustic pointers, as well as variables such as transmission energy, pitch, silence removal, windowing, Mel consistency, amplitude Cepstrum, and jitter, are used to analyze the spoken signal..At the ultimate finish, the classification strategy i.e. Support vector machines are utilized to identify the quality and pathological speech, based on the options extracted in the previous section. The Speech pathology recognition system could successfully categorise and tag the standard tone of voice and the pathological speech which contributed to diagnosing the patient.

Keywords:- *Pathological, Cepstrum, Support Vector Machine, spectrogram.*

I. INTRODUCTION

Speech plays a serious role in associate individual's participation within the society. Varied voice and hearing disorders limits the active participation of the person within the social activities.

Dysarthria may be a fibre bundle disorder that results from weakened movement of muscles employed in speaking. Dysarthric speech exhibits slow rate when put next with normative speech, however bound words inside the speech is made at a quicker rate.

The mechanism concerned in speaking includes respiration, phonation, resonance and articulation. Defect of speech causes problem in coordinating and dominant the muscles employed in the assembly of speech as well as tongue, lips and vocal folds. This ends up in poor articulation inflicting the dysarthric speech to be unintelligible inflicting the speech to be less intelligible and troublesome to know.

II. THEORITICAL BACKGROUND

An identification system for Voice Pathology was designed and enforced using Victimization SVM Classifiers and Naïve Thomas Bayes Classifiers. Pre-processing techniques like silence removal, filtering, and windowing were used. Mel-frequency Cepstrum was taken as Feature Extraction Techniques; Classification techniques adopted here are SVM & Naïve Thomas Bayes classifier. Mythical creature curve was aforethought to seek out the simplest performance among the 2 classification algorithmic program, in this, SVM has the good accuracy ninety eight whereas scrutiny with the Naïve Thomas Bayes accuracy 94%[1].

The comprehensibility of dysarthric speech is improved exploitation feature house mapping of LPC options towards developing a good medical care tool. The comprehensibility of the changed speech is measured subjectively exploitation DMOS listening check and objectively exploitation PESQ; they need determined important enhancements over the dysarthric speech with Associate in Nursing comprehensibility improvement of sixty three and forty three. 4% for DMOS and PESQ measures severally. This approach is employed in building a medical care tool that is straight forward, robust, portable. The medical care tool thus supports the patients to regain confidence to move a lot of, thereby rising their participation within the society [2].

To train the system that detects whether or not someone suffers from defect of speech or not by extracting time domain options like noise and shimmer from their speech input. The noise and shimmer options of the disordered voice samples were ascertained to be larger than healthy voice samples. This technique is often any developed to notice and classify the speech disorders like brain disorder, defect of speech and Specific Language Impairment (SLI) [3].

Using the sample corpus as an input, the speech knowledge was aforethought as waveforms, and the options were extracted from the wave files. According to observations and MFCCs, higher incapacity speeds result in lower recognition accuracy. Silence between phonemes, frames of input, where it is less than a minimum worth, is cut down and discarded as filler. The popularity accuracy of the system developed can verify the triple-crown use of the

ISSN No:-2456-2165

strategy amongst the users. The secret writing time is additionally a metric collected during this study, that is calculated because the magnitude relation of secret writing time over recording time. Suppose Associate in Nursing audio file features a recording time (RT) of three hours and also the secret writing method took half dozen hours. Then the speed is counted as 2xRT [4].

III. DESIGN AND IMPLEMENTATION

Tone of Speech pathology recognition system classifies the urged speech as traditional or pathology finalizing the tone of voice in 3 steps:

a) Preprocessing
b) Extracting the many options from the preprocessed transmission
c) Utilizing the extracted features, classifying the speech into normal or pathology

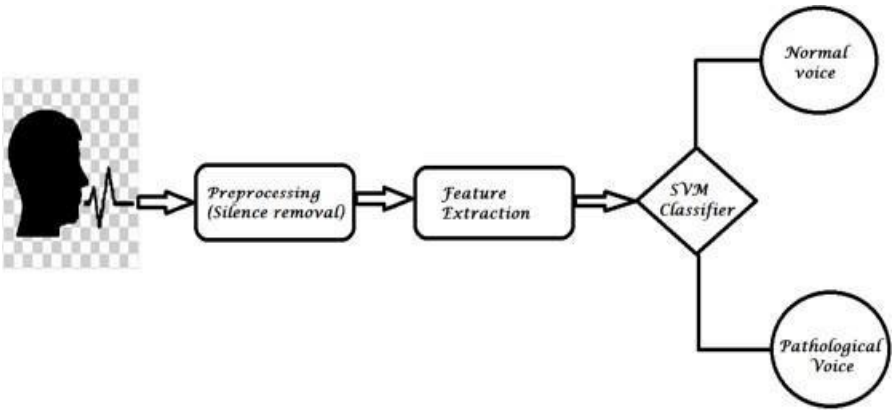

Fig. 1: Block diagram of the system

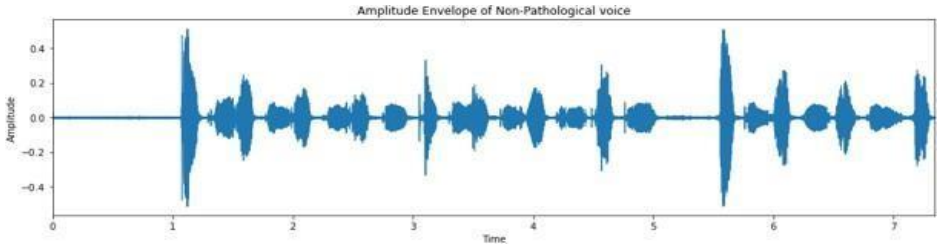

Fig. 2: Waveform of normal voice

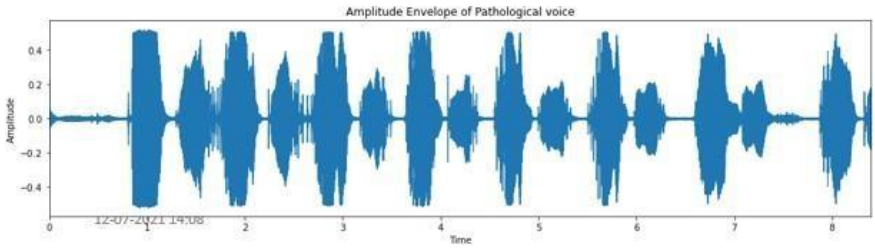

Fig. 3: Waveform of pathological voice

To understand the variation within the patterns of input, the chosen cluster of files was compared in undulation analysis and exposure read. The undulation patterns shows that a similar word was spoken at completely different length of the time and someone with a lot of impaired speech takes longer time to pronounce the word. This necessitates the requirement to compact the signal and increase the speed of vocalization.

- ***Spectrogram view***

A photograph could be a visual method of representing the signal strength, or "loudness", of a proof over time at varied frequencies gift in a very explicit wave shape. It conveys the signal strength victimization the colors – brighter the color the upper the energy of the signal.

A. Preprocessing

As a part of the preprocessing step, silence is removed from the audio tracks and the energy of your time and spectral center of mass is utilized.. Each of those options, filters the voiced examples by fitting a threshold on the speech transmission, then the tone of voice was metameric

ISSN No:-2456-2165

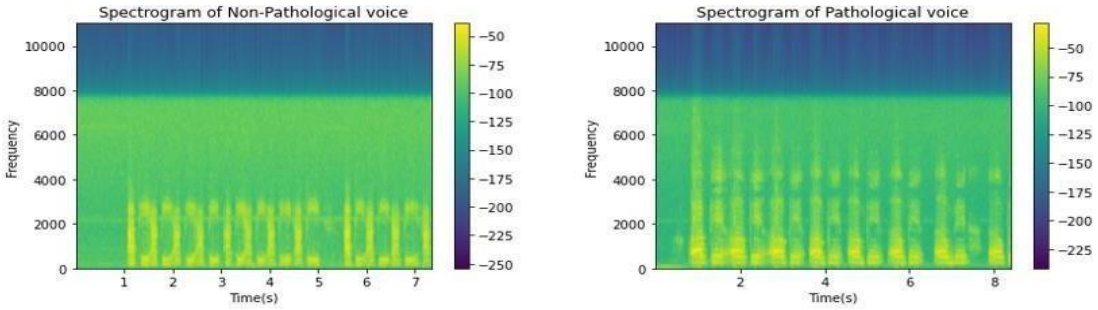

Fig. 4: Spectrogram of normal and pathological voice

and therefore the tone of voice samples lies below the brink are cared as silence and therefore the different samples are taken because the speech section. The amplitude of the tone of voice is throw in silence section and therefore the amplitude are larger if the transmission phase contains voice communication. Such amplitude deviation over associate degree interval is recognized by the short-time energy operate. Spectral center of mass could be used to urge the brightness of the sound. It is assessed by crucial the gravity i.e., the weighted mean of the frequency that are inside the voice signals, exploitation the Fourier Frequency Transformation. Spectral center of mass locates the center of the spectral energy circulation, that is additionally referred to as associate degree equilibrium purpose of the vary. A amount wherever the speech signals are taken for process is known as Window. So the tone of voice or speech information inside the Window is known as Framework.

In this paper, the area of the framework is named by its Size and also the overlapping size is named Framework switch Length. The playacting windowing approach is employed because of its larger operating half and its own smoothness within the reduced pass. Voice samples from TORGO information were taken for extraction of options. The information consists of each healthy and disordered voice samples. All the voices are recorded at 16kHz frequency and are in wave format.

Since the TORGO information was recorded for the experimentation purpose, it had been recorded in an exceedingly quiet atmosphere using a microphone array. Therefore, the noise removal and silence removal steps are skipped.

- *Noise Removal:* Suppressing the background noise that is caused by the unwanted signals.
- *Silence Removal:* To scale back the time interval and to extend the performance of system by eliminating unvoiced segments from the signaling.

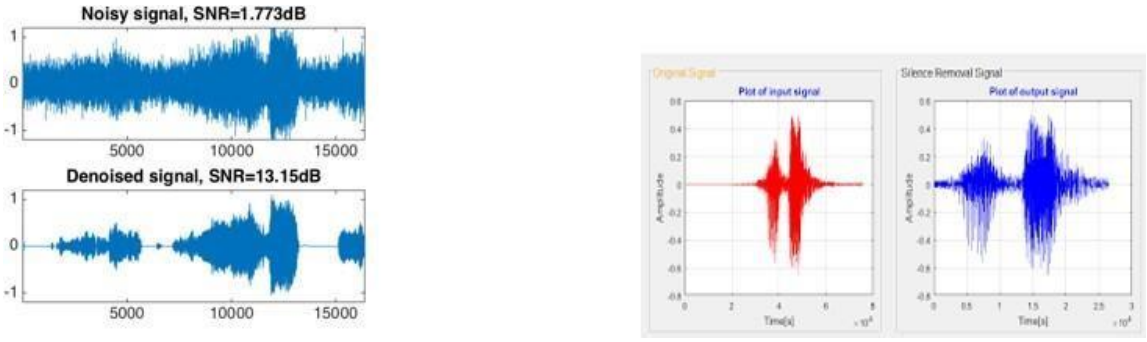

Fig. 5: Noise and Silence Removal

B. Feature Extraction

Feature Extraction performs a big role in selecting the options that is compatible for the classification method.

The speech wave is shortened to get rid of silence or physical science interference that will be gift in beginning or finish of the sound file. it's how to spice up the signal's high-frequency elements, deed the low frequency elements in their original state.

- *Framing:*Speech signal is sampled at 16kHz. every speech sample taken is fastened to a definite length therefore on acquire equal range of frames from each sample with rate constant.
- *Hamming Windowing:*Discourse flag is ceaselessly shifted each in time and recurrence. So, the discourse flag is analyzed not as a whole in any case it's expected that this discourse flag is Quasistationary in nature. Hence, the short time analysis of the signal is completed by the method of windowing. it's wont to smoothen the top values of the truncated signal obtained within the previous stage. The signal once framing is split into overlapping frames of N samples every i.e the frame size is N. If adjusted frames ar separated by M samples wherever N>M, then the acting window equation is given by

ISSN No:-2456-2165

$$w(n) = \left[0.54 - 0.46\cos\left(\frac{2\pi n}{M-1}\right)\right]$$

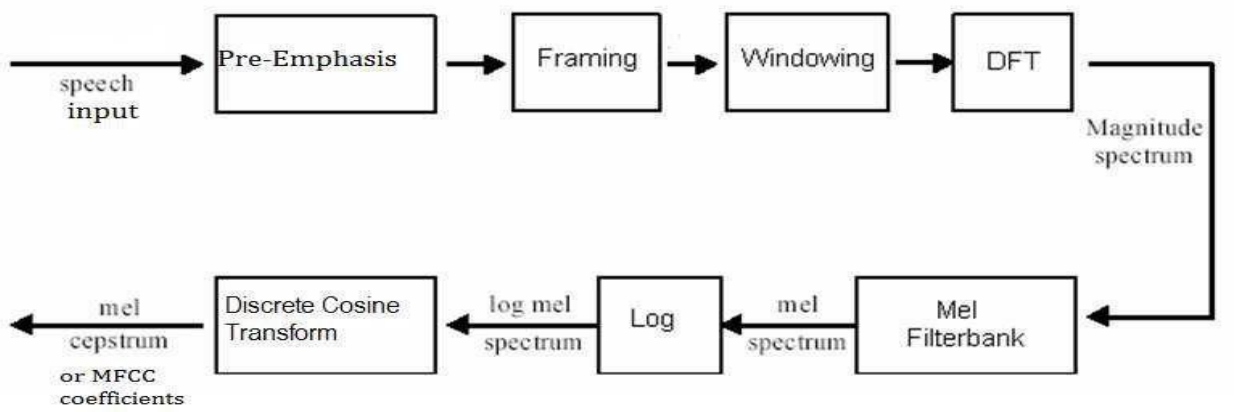

Fig. 6: Mel Frequency Cepstral Coefficients

MFCC i.e., Mel Frequency Cepstral Coefficients, is one in all the simplest feature extraction techniques.

In speech analysis, Mel Rate of prevalence Cepstrum could be an example of an influence spectral vary of a voice transmission, that made and predicated on the linear circular function amendment of the log power vary over a nonlinear Mel-scale rate of prevalence.

The Mel Frequency Cepstrum uses equally spaced frequency bands on the Mel Scale, which estimates the human exteroception system's response additional closely than the quality Cepstrum that uses linearly-spaced frequency bands.

C. Pre Emphasis:

- *Discrete Fourier Transform:*This block converts the speech data from the time domain to the frequency domain. The FFT method is applied to each frame in order to calculate the DFT coefficients.

- *Mel Filter Bank and DCT:*

The spectrum obtained after DFT is converted to Mel frequency by Mel Filter Bank and Mel scale I is used Mel–scale is a logarithm–based scale and it aligns with the human perception of sound intensity, which is measured in decibels(dB).The Mel Frequency Cepstrum Coefficients use frequency bands on the Mel Scale equally.

For a given frequency f, calculation of the Mel is given by below equation

$$m = 2595\log_{10}\left(1 + \frac{f}{700}\right) = 1127\ln\left(1 + \frac{f}{700}\right)$$

The cepstrum provides an honest illustration of the speech signal that is that the key for representing and recognizing characteristics of the speech information. The resulted power spectrum is filtered victimisation this triangular filter bank created with Mel-scale. Then the coefficients will be obtained from the filtered spectrum by taking the power of sub band energies followed by the separate circular function Transform(DCT).

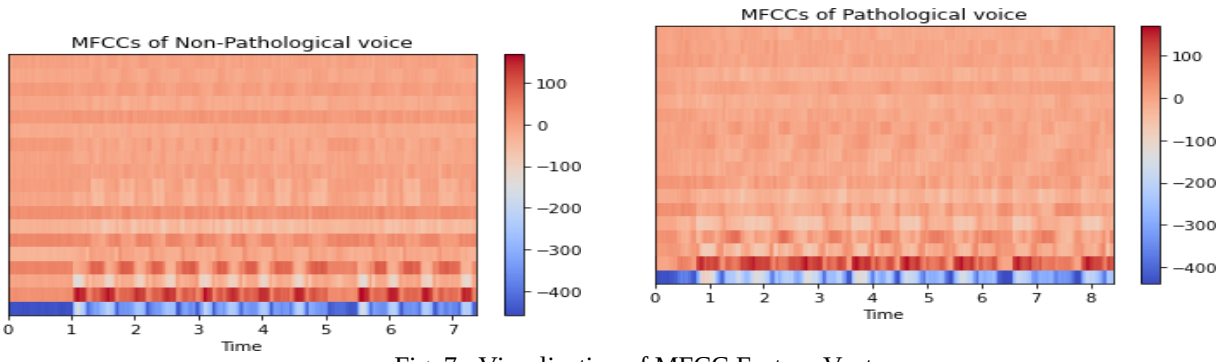

Fig. 7: Visualization of MFCC Feature Vector

ISSN No:-2456-2165

Plot of Mel Frequency Cepstrum Coefficients of Normal and disordered voice samples are shown in the below figures:

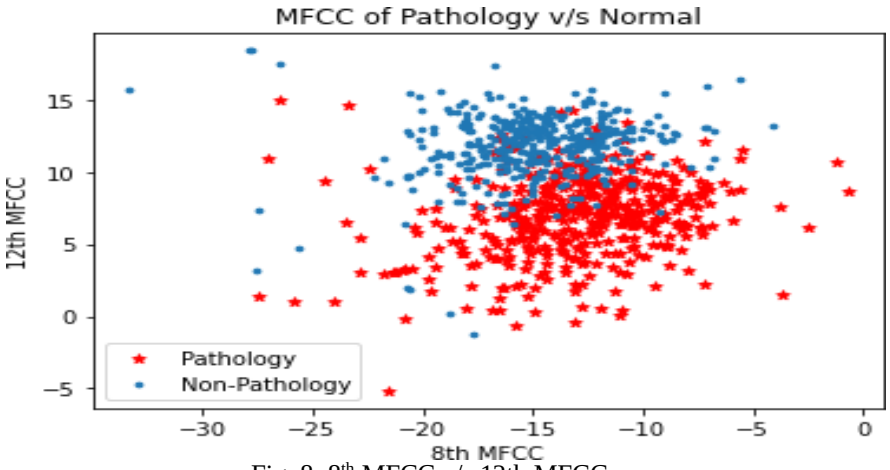

Fig. 8: 8th MFCC v/s 12th MFCC

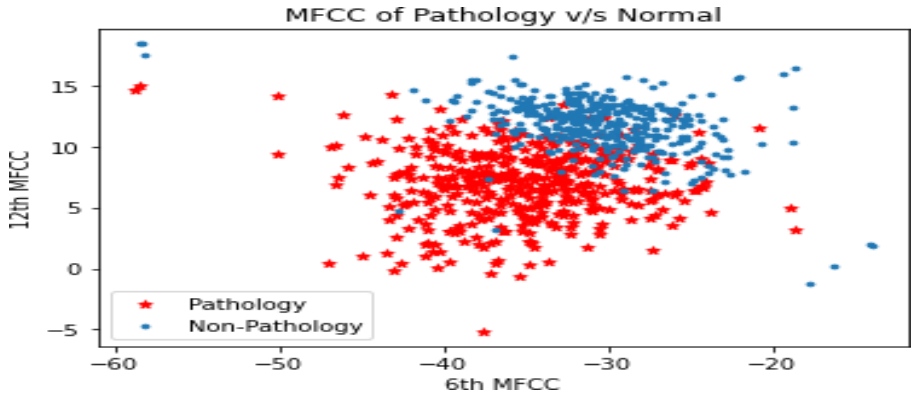

Figure 9:6th MFCC v/s12thMFCC

- SVM Classification: Support Vector Machine is one of the foremost common supervised Learning algorithms, that's utilized for Classification what is more as Regression problems. However, primarily, it's used for Classification problems in Machine Learning. The goal of the SVM algorithm is to make the only line or decision boundary that will segregate n-dimensional space into classes so as that we are going to merely place the new information at intervals the right category at intervals the long run.

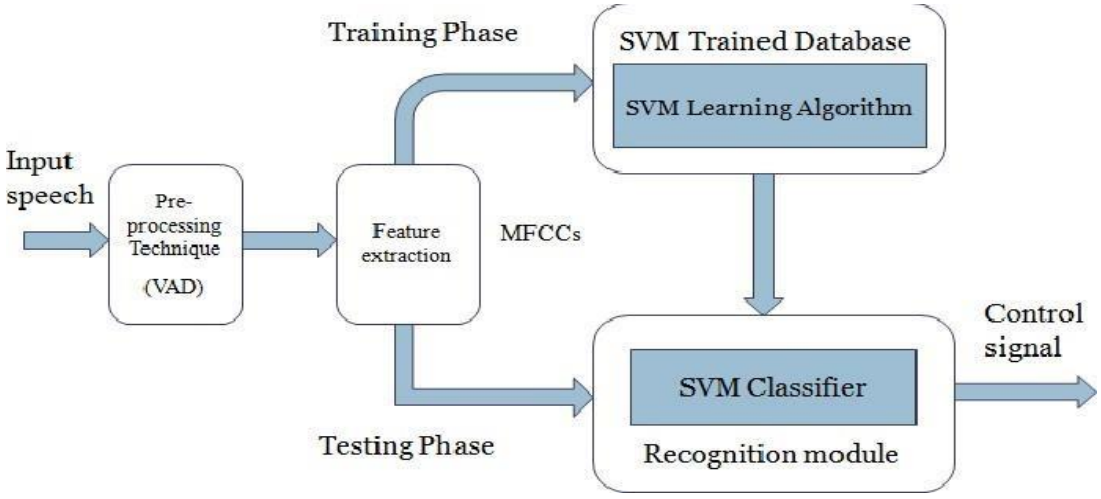

Fig. 10: Classification using SVM

ISSN No:-2456-2165

Here the best decision boundary is termed as hyper plane. SVM chooses the acute points/vectors that facilitate in creating the hyper plane. These extreme cases square measure called as support vectors, and thus formula is termed as Support Vector Machine. The Separators Wide Margin or Support Vector Machines (SVM) square measure wide utilised in applied math learning techniques, and it had nice success within the majority areas where they were applied. In speech recognition, SVM had reached a high accuracy rate larger than ninetieth in most of the researches. Hyper planes square measure the selection boundaries that facilitate to classify the data points classes of SVM samples. Also, the dimension of the hyper plane depends upon the quantity of options. They will be negative or positive in nature.

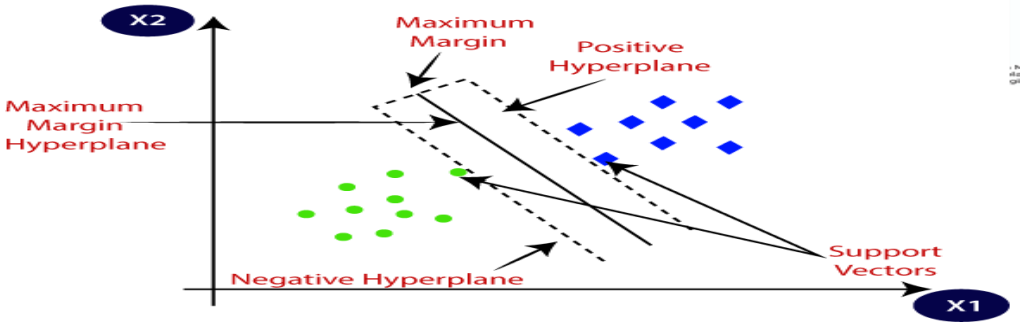

Fig. 11: SVM Hyperplane visualization

Data points falling on either facet of the hyper plane are attributed to altogether completely different classes of SVM samples. Also, the dimension of the hyper plane depends upon the number of choices. They will be negative or positive in nature.

The data set is labeled , for i= 0….N, xi Rd and Loloish, where xi is that the pc file and Loloish is that the corresponding class. SVM use many kernels to map data file into high-dimensional choices space like linear, sigmoid, polynomial, radial basis function (RBF).

• *SVM kernel Functions:*

Kernel function	Expression	Parameter
Liner kernel function	$K(x_i, x_j) = x_i \cdot x_j$	
Polynomial kernel function	$K(x_i, x_j) = (x_i \cdot x_j + 1)^d$	d
Radial basis function (RBF) kernel function	$K(x_i, x_j) = \exp\left(-\gamma \left\|x_i - x_j\right\|^2\right)$	$\gamma > 0$
Sigmoid kernel function	$K(x_i, x_j) = \tanh\left(b(x_i, x_j) + c\right)$	b, c

Table1:SVM Kernels

SVM algorithms use a bunch of mathematical functions that area unit referred to as kernels. The perform of a kernel is to want info as input and rework it into the specified kind. altogether totally different SVM algorithms use differing kinds of kernel functions. These functions area unit of assorted kinds for instance, linear, nonlinear, polynomial, radial basis perform (RBF), and sigmoid. the foremost most well-liked quite kernel perform is RBF. as a results of it's localized and contains a finite response on the entire axis. The kernel functions return the important between two points in Associate in Nursing extraordinarily acceptable feature house. thus by method a notion of similarity, with a touch computing worth even among the case of very high-dimensional areas.

$$K\left(\overline{x}\right) = \begin{matrix} 1 & \text{if} \left\|\overline{x}\right\| \le 1 \\ 0 & \text{otherwise} \end{matrix}$$

ISSN No:-2456-2165

IV. RESULT AND DISCUSSIONS

The audio data has been taken from TORGO database for pathological and normal voices in .wav format. To develop the program Python language is used in Jupyter notebook software. Librosa library is imported in order to work with audio files. Many other libraries which are required are also imported. The audio files are loaded and converted into numerical form for further analysis using one of the librosa functions. Toextract the Mel Frequency Cepstral coefficients, result from the previous step is used.

Here 19 coefficients/ features were extracted. MFCC feature vector is then converted into data frame for classification purpose. All the above mentioned steps are applied for both Pathological and Non- Pathological voice samples. For the pathological data frame a target column is created with value 1 (To indicate as Pathology).Similarly for the Non-pathological data frame a target column is created with value 0 (To indicate as Non-Pathology). Then the two data frames are combined/concatenated resulting a single data frame.

To train the Machine learning model below steps are followed:

- The target column is separated and dropped from the data frame.
- The 5 ML rules are satisfied.
- The data is split into training and testing data set.
- Then the classification model is trained with SVM algorithm using training dataset.

Score is evaluated using testing dataset which approximately equal to 98%. Other classification algorithm such as KNN, Decision tree Classifier & Random forest classifier were also employed, among them SVM is selected due its comparatively high prediction and accuracy. This can also be verified by plotting confusion matrix.

a) Confusion matrix for SVM

b) Confusion matrix for KNN

A confusion matrix (error matrix), may be a specific table layout that enables visual image of the performance of associate degree rule

c) Confusion matrix for Decision tree classifier

d) Confusion matrix for Random Tree forest

Fig. 12: Confusion matrix for SVM, KNN

A. ***Random Tree Classifier and for Decision Tree Classifier respectively***

It describes the performance of a classification model on a group of take a look at information that verity values square measure legendary. Currently the model is prepared to classify the voices. The opposite external samples square measure evaluated exploitation the model.

Fig. 13: GUI for classification of voices

V. CONCLUSION AND FUTURE SCOPE

The info of varied voice samples of pathological and non-pathological person of each gender is collected/downloaded from TORGO info. Here Python code is employed for the implementation purpose. The techniques used for Pre-processing are silence removal; filtering and windowing were processed and obtainable in TORGO info already. Mel-frequency Cepstrum is taken as Feature Extraction Technique and co-efficient square measure obtained. Classification technique SVM is compared with different classifiers, SVM provides higher performance than different classifiers. In future this technique are often additional increased to sight and classify additional speech disorders like brain disorder, defect of speech and Specific Language Impairment (SLI).Also, the quantity of severity of the disorder may be projected.

REFERENCES

[1.] N A SheelaSelvakumari and V Radha "A Voice Activity Detector using SVM and Naïve Bayes classification algorithm", IEEE paper, 2017.
[2.] SreejuSivaram, C Santhosh Kumar and A Anand Kumar "Enhancement of Dysarthric Speech for Developing an Effective Speech Therapy Tool", IEEEWISPNET 2017.
[3.] Ashutosh Singh, AbhishekKittur, KalpeshSonawane, Ayushman Singh and Dr.SavithaUpadhya "Analysis of Time Domain Features of Dysarthria Speech", IEEE Xplore ICCMC 2020.
[4.] Balaji V and Dr. G Sadashivappa, "Waveform Analysis and Feature Extraction from Speech Data of Dysarthric persons", IEEE SPIN2019.
[5.] ImenHammami, LotfiSalhi and Salam Labidi "Pathological Voices Detection using Support Vector Machine",2nd Internal Conference onATSIP(2016)..
[6.] T.Orzechowski, A.Izworski, R.adeusiewicz, K.Chmurzyn'ska, P.Radkowski, L.Gatkowska and

Krakow, Poland "Processing of Pathological changes in Speech caused by Dysarthria".

[7.] Ahmed Alnasheri, Ghulam Muhammad, Mansour Alsulaiman, Zulfiqar Ali, Khalid H. Malki, Tamer A.Mesallam, and MohamedFarahat "Pathology Detection and Classification using Auto-correlation and entropyfeatures in Different Frequency Regions",IEEE Access, Vol.6, 2017

[8.] Sudarsana Reddy Kadiri and PaavoAlku, Senior Member,IEEE. "Analysis and Detection of Pathological Voice using Glottal Source Features", IEEE Journal Of Selected Topics In Signal Processing (Volume: 14, Issue: 2, Feb. 2020.