Development of Automation System for Disease Disorder Diagnosis using Artificial Neural Networks and Support Vector Machine

Syed Mohammad Ali and Dr. Pradeep Tulshiram Karule

Abstract--- *Voice disorder classifications have developed more and more momentum now days because of complication in conventional methods. speech disorder diseases creates voice problem hence speech signal can work as useful tool to diagnose voice disorders.*

In this research work, normal & abnormal speech signals are taken & a system is designed to classify patients of chordectomy, laryngitis, laryngeal paralysis, psychogenic dysphonia, vocal cord cancer from normal. Speech signals are first preprocessed. The preprocessed signal is used for spectral analysis with which normal & abnormal speech signals are differentiated.Various features are extracted and after selecting relevant and efficient feature, these features are given to various artificial neural networks and SVM. The neural network used are MLP,GFF,Modular network. The accuracy for these networks is 60.3,65.97,52.57 respectively .Support vector machine was found to be optimum classifier with classification of 92.26% and the percentage correct dtermination for various disease like chordectomy, laryngitis, laryngeal paralysis, psychogenic dysphonia, vocal cord cancer and normal using SVM is 100, 100 , 94.28, 95.34, 81.25 and 84.9 respectively

Index Terms--- *Chordectomy, Laryngitis, Laryngeal Paralysis, Psychogenic Dysphonia, Vocal Cord Cancer, Pathological Speech Signals, Preprocessing, Spectral Analysis, Feature Extraction, Support Vector Machine*

Syed Mohammad Ali, Prof., Department of Electronics & Telecommunication Engineering, Anjuman College of Engineering & Technology, Nagpur, India.

Dr. Pradeep Tulshiram Karule, Department of Electronics Engineering, Yeshwantrao Chavan College of Engineering, Wanadongri, Nagpur, India

I. INTRODUCTION

Automation system in voice disorder detection has received great momentum recently. Digital signal processing has turn out to be a significant means for voice disorder recognition[3]. The causes of psychogenic dysphonia can be very unusual.The various causes can be mental stress due to unsettled grief ,tense situations, harassment, untreated fear and shock experiences, unconscious conflicts ,neuroses & depression. Older patients are affected by vocal cord cancer. The patient suffering from vocal cord carcinoma leads to early hoarseness . Laryngitis is an inflammation of the larynx. It causes hoarse voice ,temporary or complete loss of the voice because of irritation to the vocal folds. Dysphonia is the medical term for a vocal disorder, and laryngitis is one reason. In vocal cord paralysis, the nerve impulses to your voice box (larynx) are damaged, resulting in paralysis of the muscle. Even doctors may not be knowing the cause of vocal cord paralysis.The Known causes may include:Injury to the vocal cord during surgery, Neck or chest injury, Stroke, Tumors, Inflammation, Neurological conditions**.** If you have definite neurological conditions, such as multiple sclerosis or Parkinson's disease, you may experience vocal cord paralysis. Chordectomy is the surgical removal of a cord. It usually refers to removal of the vocal cord, often for the purpose of treating Laryngeal Cancer. Pathological voice signal of above disease & Normal signals are taken from German voice disorder database in which the patients have pronounced vowel like ‘a’ [19].The sampling frequency is 50khz and the bit rate is 800 kbps .It is possible to diagnose disease

using certain feature of speech signal [3].Speech signal in non-intrusive in nature& it has potential for providing quantitative data with reasonable analysis time. Due to nature of jobs, unhealthy social habits people are subjected to risk of voice problem [9].So study of speech signal of pathological voice has become an important topic for research as it reduces work load in diagnoses of pathological voices [8]. The algorithm shown in figure 1.below shows the flow of control. Here in this paper we have taken speech samples of specified disorder and normal persons. These speech samples are passed through and high pass filter, the filtered output is framed and then each frame is passed through window. The output signal which is framed and windowed is used for spectral analysis. In spectral analysis derivative of logarithmic spectrum is taken, then logarithmic spectrum is used to get cepstrum. Framed signal is also used for finding autocorrelation of speech signal. From spectrum, cepstrum and autocorrelation, pattern classification is done for finding normal and pathological signal.

Speech signal is sinusoidal in nature with different frequency, different amplitude, & different phase [6]. Speech production requires close support of numerous organs which from the phonetic point of view these organs may be divided into Lungs, Bronchi, Tracheas (producing expiration air steam necessary for phonation) Larynx (amplifying the initial tone) .Root of the tongue, throat, nasal cavity, oral cavity (forming tone quality & speech sound) [7].The use of noninvasive techniques to evaluate the larynx and vocal tract helps the speech specialists to perform accurate diagnose [10].The figure 1 shown below is algorithm of pathological and normal speech classification.

II. ALGORITHM FOR CLASSIFICATION OF PATHOLOGICAL SPEECH

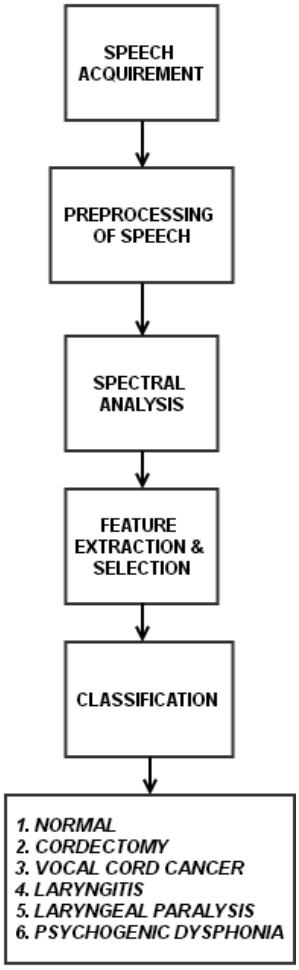


Fig. 1: Algorithm of Pathological Speech Classification

III. PREPROCESSING OF SPEECH

Preprocessing of speech involves passing speech signal through, pre-emphasis filter and then making frames of the speech of 20 to 100 msec. This framed speech is passed through window as window smoothens the signal.Design of this filters is given below.

A. Pre-Emphasis Filter Design

The input to pre-emphasis filter is noisy signal. The pre-emphasis filter is high pass filter is used to flatten the speech signal spectrum & to make the speech signal less sensitive to finite processing effects later in speech signal processing [4]. The pre-emphasis filter amplifies the area of spectrum. Thus improving the efficiency of spectral

analysis [2]. The time domain presentation of filter is given by[5].

$$Y(n) = X(n) - \lambda X(n-1) \quad (1)$$

Where y (n) is the output, x (n) is input speech sample & λ is the filter coefficient with λ = 0.9375 optimum result of filtering is received. The output if this filter is framed &passed through window this is done as speech signals are analyzed for short period of time (5 msec to 100msec). The signal is fairly stationary & windowing is done to avoid problem due to truncation of signal & window helps in smoothening of signal [1].This filter designed can be used for preprocessing of any automatic speech recognitions systems.

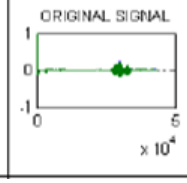
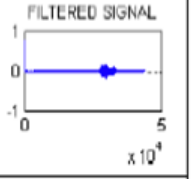
Sr. No.	Patient Samples	Disease	HP Filter Output
1		Squamous Cell Carcinoma	

Fig. 2: Result of Filtering of Speech Signal

IV. CLASSIFICATION OF SPEECH USING SPECTRAL ANALYSIS AND AUTOCORRELATION

A. *Cepstrum and Derivative of Spectrum*

'cepstrum' word has come from first four letter of spectrum [13]. This is a reliable way of obtaining an estimate of the dominant fundamental frequency for long clean stationary speech signal. The cepstrum is a Fourier analysis of the logarithmic amplitude spectrum of the signal. If the log amplitude spectrum contains many regularly spaced harmonics, then Fourier analysis of the spectrum will show a peak corresponding to the spacing between the harmonics i.e. fundamental frequency. Here signal spectrum is treated as another signal, then looking for periodicity in the spectrum itself. The cepstrum is so called because it turns the spectrum inside out. The X axis of cepstrum has unit of quefrency & peak in cepstrum is called rahmonics [12].If X(n) is the speech signal then logarithmic spectrum is given by

$$Y(n) = FFT[X(n)] \quad (2)$$

$$Y(n) = 20 \times \log_{10}[abs\, Y(n)] \quad (3)$$

The cepstrum is DFT of log spectrum

$$Y(n) = FFT\left[\log\left(abs\,(Y(n))\right)\right. \quad (4)$$

Figure 3 & 4 shows classification of diseased and normal speech using cepstrum.

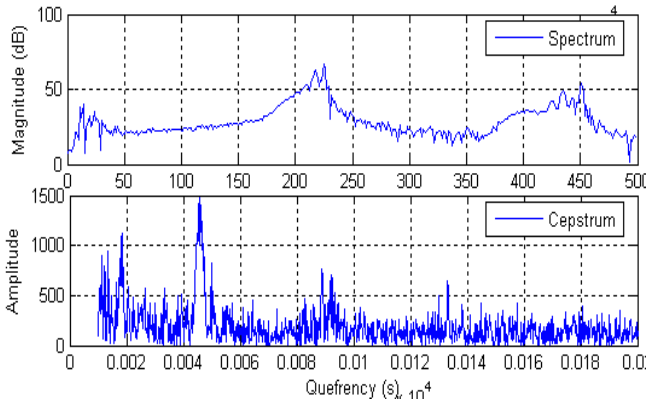


Fig. 3: Cepstrum and Spectrum of Vocal Cord Cancer Normal Person

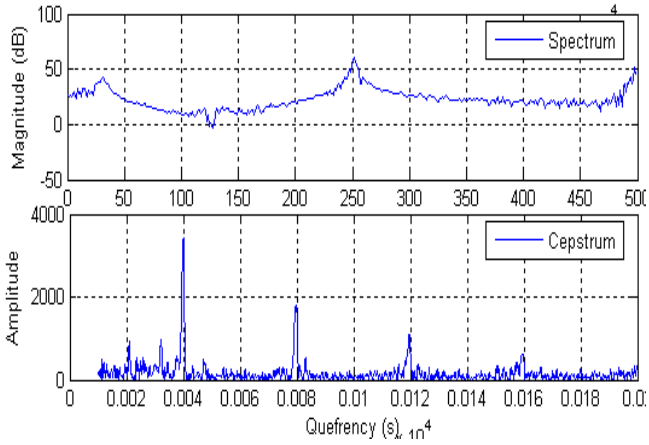


Fig. 4: Cepstrum and Spectrum of Normal Person

B. *Auto Correlation of Pathological and Normal Signal*

One of the other time domain methods, which is applied for the classification of pathological speech signal from normal is Autocorrelation method. Using this method, one can easily classify the normal and vocal cord cancer patient speech signals. The autocorrelation of discrete time signal

X(n) , is given by [6].

$$r_{xx}(l) = \sum_{n=-\infty}^{\infty} X(n).X(n-l) \quad l = 0, \pm 1, \pm 2, \ldots \quad (5)$$

The autocorrelation function of a signal is a transformation of signal, which is useful for displaying structure in the waveform [11]. For the normal signal the decay of autocorrelation of signal with respect to time is exponential whereas for abnormal decay will not be exponential.Maximum autocorrelation was used as one of the features.

C. Spectrogram

Another distinct signal processing techniques used for voice analys is, the spectrogram .is commonly used as it allow for visualization of variation of energy, of the signal as function of both time and frequency [14]. The study investigates the use of the global energy of the signal estimated through spectrogram as a tool for discrimination between signals obtained from healthy and pathological subjects. The vertical axis of the spectrogram is frequency and it provides an analysis of signal into different frequency regions as shown in figure 5 & 6.You can think of each of these signals as comprising a particular kind of building blocks of signal. If a building block is present in the signal at particular time then highlighted region will be shown at the frequency of building block and time of the event [15].

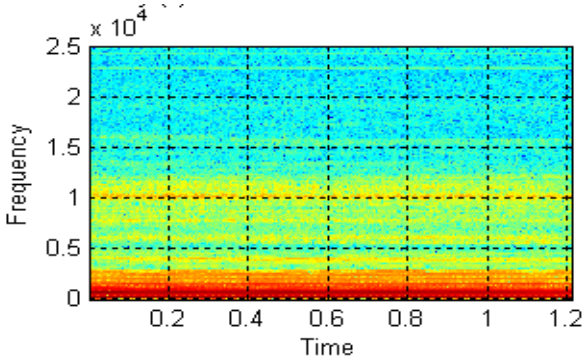


Fig.5: Spectrogram of Normal Person Time

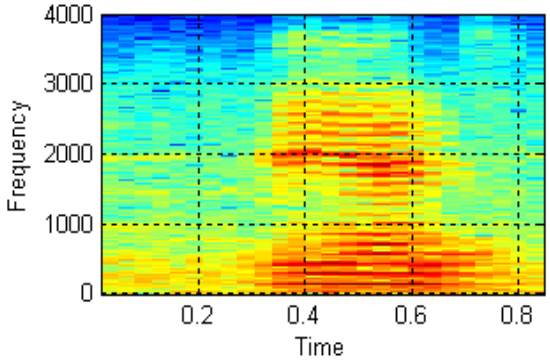


Fig.6: Spectrogram of Vocal Cord Cancer Patient

V. FEATURE EXTRACTION FEATURE SELECTION

A. Pitch Calculation

Dominant frequency of speech signal is called pitch. Pitch detection task an essential task in a variety of speech processing applications. Although many pitch detection algorithm both in time and frequency domains, have been proposed [16]. However, performance improvement in noisy environments is still desired [17].Here we are proposing cepstrum method, which shows good performance for quasi-periodic signals [18]. And manual method to get the pitch of normal and pathological signal. The Figure 7&8 shows the calculation of pitch using proposed method of cepstrum. As we know that x-axis of cepstrum has unit of quefrency & peaks in cepstrum (which relates the periodicity in the spectrum) are called rahmonics. To obtain an estimate of the pitch from the cepstrum we look for the peak in the quefrency region corresponding to typical speech fundamental frequencies (1/quefrency). The pitch of the signal under consideration is found to be 129.668Hz.

The second method which is proposed is manual method here the period of the segment can be calculated by finding the time difference of two successive peaks of figure 9 and 10 and doing the calculation of pitch by manual method.

$$F = \frac{1}{T} = \frac{1}{T_2 - T_1} = \frac{1}{0.054 - 0.0463} = 129.87\ Hz \quad (6)$$

$$F = \frac{1}{T} = \frac{1}{T_2 - T_1} = \frac{1}{0.05 - 0.04166} = 119.33\ Hz \quad (7)$$

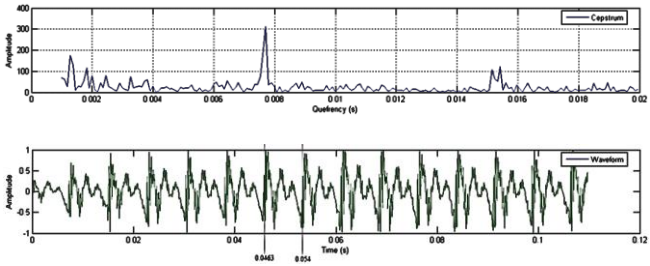


Fig. 7: Calculation of Pitch using Cepstrum & Manual Method

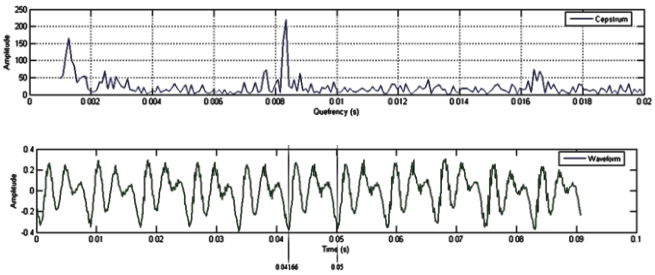


Fig. 8: Calculation of Pitch using Cepstrum & Manual Method

B. *Formant Frequency Estimation*

Formants are frequencies of resonance for each frame. It is often measured as an amplitude peak in frequency spectrum of the speech formants are resonances of the vocal tract [20].The formant frequency are calculated using linear predictive coding (LPC). The formant frequency is obtained by finding the roots of prediction polynomial. The LPC finds the best IIR filter from the section of speech signal and then plots frequency response of filter as shown in figure9 .The system function and difference equation of LPC filter is given below.

$$H(z) = \frac{X(z}{E(z)} = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}} = \frac{1}{A(z)} \quad (8)$$

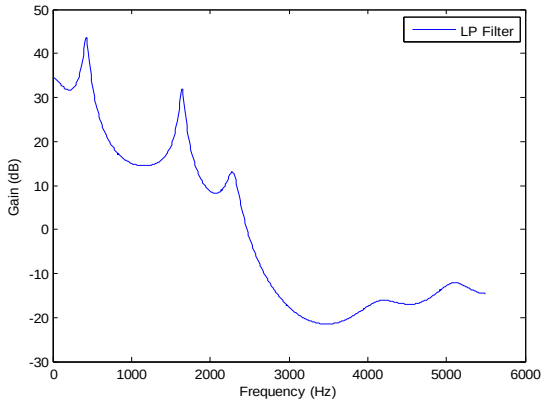


Fig. 9: Frequency Response of LPC Filter Showing Formants

$$x[n] = \sum_{k=1}^{p} a_k x[n-k] + e[n] \quad (9)$$

Formant 1 Frequency 426.4 Hz

Formant 2 Frequency 1205.1 Hz

Formant 3 Frequency 1643.3 Hz

Formant 4 Frequency 2293.7 Hz

Formant 5 Frequency 4151.0 Hz

Formant 6 Frequency 5088.7 Hz

C. *Short Time Energy (STE)*

Pathological speech signal has greater amplitude variations as compared normal speech; here short time energy has been used to counter the amplitude variation in pathological speech signal. The short time energy is expressed as

$$E_{\hat{n}} = \sum_{m=-\infty}^{\infty} (x[m]w[\hat{n} - m])^2 = \sum_{m=-\infty}^{\infty} x^2[m]w^2[\hat{n} - m] \quad (10)$$

D. *Short time Zero Crossing Rate (ZCR)*

Short time ZCR is defined as the number of times the speech signal changes sign within a given window

The ZCR in case of stationary signal is defined as,

$$Z_{\hat{n}} = \sum_{m=-\infty}^{\infty} 0.5\ |sgn\{x[m]\} - sgn\{x[m-1]\}|w[\hat{n} - m] \quad (11)$$

$$sgn\{x\} = \begin{matrix} 1 & x \geq 0 \\ -1 & x < 0 \end{matrix}$$

If the number of ZCR is more, then signal is changing rapidly then the signal may contain high frequency information & vice versa [23, 24].

E. Spectral Centroid (SC)

The spectral centroid measures the "brightness of sound", i.e. it measures where most of the power in the speech segment is located .it has been previously used for detection of clinical depression. It is the weighted average of the frequency of the spectrum and thus would give us an idea as to what frequency range most of the power of spectrum would lie in. We wanted to analyze if there is shift in the spectral centroid for pathological voice toward higher frequency. The spectral centroid is given by

$$SC = \sum_{k=0}^{N-1} X(k)F(k) \,/\, \sum_{k=0}^{N-1} X(k) \quad (12)$$

Where X(k) represents weighted frequency value or magnitude of bin number k, F(k)represents center frequency of that bin[25].

F. Spectral Flux (SF)

The spectral flux is defined as difference in the power spectra of two consecutive speech frames. Thus, it measures the frame-to-frame variability in the spectral shape [23, 24]. Spectral flux, also called Delta Spectrum Magnitude, is a measure of the rate of change of the spectral shape and is given by the sum across one analysis window of the square difference between the magnitude spectra corresponding to successive frames of the STFT:

$$\mathrm{F_r} = \sum_{\mathrm{k}=1}^{\mathrm{N}/2} \mathrm{X_r[k]} - |\mathrm{X_{r-1}[k]}|)^2 \quad (13)$$

G. Mel Frequency Cepstral Coefficient (MFCC)

Steps At A Glance

Frame the signal into short frames. For each frame calculate the periodogram estimate of power spectrum. Apply the Mel filter bank to power spectra, sum the energy in each filter.

Take the logarithm of all filter bank energies, Take the DCT of log filter bank energies, and Keep 12-13 DCT coefficients [21].

DFT for i^{th} frame is given by

$$S_i(k) = \sum_{n=1}^{N} s_i(n)h(n)e^{-j2\pi kn/N} \quad (14)$$

$$1 \le k \le K$$

Where h(n) is hamming window and k is the length of DFT

Power spectrum of i^{th} frame is given by

$$P_i(k) = \frac{1}{N} S_i(k)|^2 \quad (15)$$

Now we create our filter banks .The formula for creating filter bank is given below

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \dfrac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \le k \le f(m) \\ \dfrac{f(m+1) - k}{f(m+1) - f(m)} & f(m) \le k \le f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad (16)$$

Where M is number of Mel filters we want f() is the list of M+2 Mel spaced frequencies. One of these real-valued transforms is the Discrete Cosine Transform (DCT), in which the basic functions are cosines. The N-point DCT of a sequence of N samples is defined as:

$$X[k] = \sqrt{\frac{2}{N}}\alpha[n] \sum_{n=0}^{N-1} x[n] \cos\left(\frac{\pi k(2n+1)}{2N}\right), \quad (17)$$

$$0 \le k \le N-1$$

Where, $\alpha[n]$ is given by

$$\alpha[n] = \begin{cases} 1/\sqrt{2} if & k = 0 \\ 1 \quad if 1 \le k \le N-1 \end{cases}$$

The DCT has the property that its energy is concentrated in its lowest coefficients.For this reason, the DCT has found its main application in the compression of data. Figure10 shows block diagram to get MFCC.

Time Envelope

Another possibility for the above mentioned smoothing is to take the maximum of the absolute amplitude values in each frame. The result is an envelope trajectory that lies on the peaks of the time-domain signal:

$$ENV_r = \max\{| Xr[n] |\} \quad (18)$$

$n = 1, N$

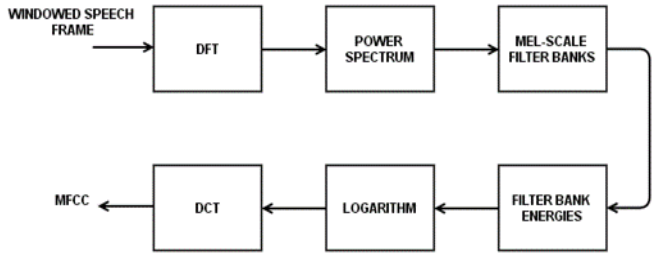


Fig. 10: Block Diagram to find MFCC

H. Root Mean Square

To obtain description of the dynamic tendency, rather than of the instantaneous amplitude of the samples. One possible way to perform this smoothing is to compute the RMS Energy of the signal in each frame, where, in this case,N is the number of samples in each analysis window.

$$RMS = \sqrt{\frac{1}{N+1}\sum_{n=0}^{N}|x[n]|^2} \quad (19)$$

I. Spectral Crest

Spectral crest is defined as ratio of peak value to RMS value of speech signal in each frame.

J. Roll Off

Here, it is defined as the frequency below which 85% of the accumulated magnitudes of the spectrum is concentrated [24]. That is, if K is the bin that

$$\sum_{k=0}^{M}|X_r(k)| = .85 \sum_{k=1}^{N/2}|X_r(k) \quad (20)$$

Fulfils then, the role off is Rr = f[K].

K. Spectral Decrease

Spectral decrease estimates the steepness of the decrease of spectral envelop over frequency given by

$$V_{SD}(n) = \frac{\sum_{k=1}^{k/2-1}\frac{1}{k}\ |X(k,n| - |X(0,n)}{\sum_{k=1}^{k/2-1}|X(k,n)} \quad (21)$$

L. Entropies

The entropy E must be an additive cost function such that $E(0) = 0$

$$E(s) = \sum_i E(s_i) \quad (22)$$

The (non-normalized) Shannon entropy

$$E1(s_i) = s_i^2 \log(s_i^2)\ so$$

$$E1(s) = \sum_i s_i|^P = \|s\|_p^p \quad (23)$$

The log energy entropy

$$E3(s_i) = \log(s_i^2)\ so$$

$$E3(s) = \sum_i \log(s_i^2) \quad (24)$$

With convention $\log(0) = 0$

The threshold entropy$E4(s_i) = 1$

If$|s_i| > p$ and 0 else where so

$E4(s) = \#\{isuchthat|s_i| > p\}$is the number of time instants when the signal is greater than a threshold ρ[22].

The "SURE" entropy.

$$E5(s) = n - \#\{isuchthat|s_i| \le p\} + \sum_i \min(s_i^2.p^2 \quad (26)$$

VI. FEATURE SELECTION

A feature should take like values within a given class, but very dissimilar values across classes.Each new feature should add as new information about the signal as possible.In this work how these features vary with respect to above disease is shown in figure 11 and if there is variation then they were considered.

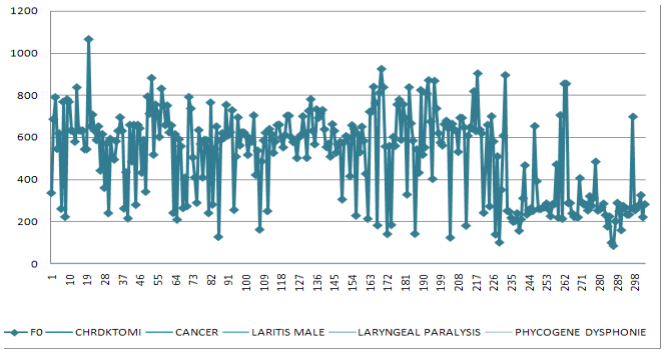


Fig. 11: Variation of Feature fo with Respect to Various Disease

VII. CLASSIFICATION

We have found out twentyfive features .The list of these features is given in the table I.The total samples taken were 776 for five disease and normal speech.Features extacted for these sample were randomized and they were applied to neural networks and SVM. Fig. 12. Shows the classification and correct determination of a disease for first 10 samples using SVM

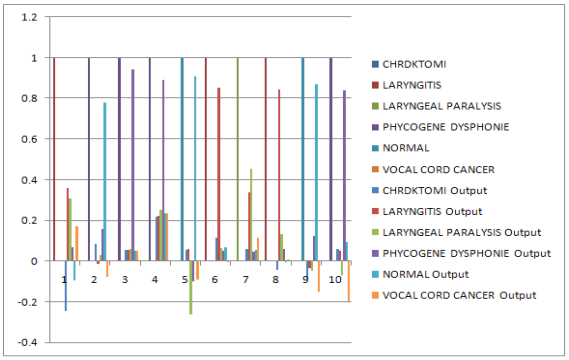


Fig. 12: Table Showing Classification and Correct Determination of a Disease for first 10 Samples using SVM

Table I: Feature Set Showing Twenty-five Features

Sr. No.	Feature set	Feature vector
1.	Mel frequency Cepstrum coefficients (MFCC)	01
2.	LPCC	01
3.	Formant FrequenciesFo,F1,F2,F3	04
4.	Pitch	01
5.	Lpc	01
6.	Spectral flux	01
7.	Spectral Centroid	01
8	Spectral Decrease	01
9.	Spectral Crest	01
10	Spectral Roll off	01
11	Entropies	05
12.	Short Time Energy	01
13.	ZCR	01
14	Peak value	01
15	RMS value	01
16	Max Autocorrelation	01
17	Standard deviation	01
18	Variance	01
	Total	25

Table II: Showing % Classification of Disease Using Various Neural Network and SVM

Network,Layers,epochs	Chordectomy	Laryngitis	Laryngeal Paralysis	Phycogene Dysphonie	Normal	Vocal Cord Cancer
MLP, 2 ,2k	40	67.44	16.12	88.88	73.80	58.82
GFF, 2 ,2k	60	72.72	47.05	77.41	87.23	26.31
MODULAR,2,2k	73.68	48.07	31.42	53.57	80.48	21.05
SVM,-,2100	100	100	94.28	95.34	84.9	81.25

Table III: Confusion Matrix and Performance Table for SVM with Accuracy 92.26%

Output/ Desired	Cordecttomy	Laryngitis	LaryngealParalysis	PhycogeneDysphonie	Normal	Vocal Cord Cancer
Cordect-tomy	22	0	0	0	0	3
Laryngitis	2	25	2	0	0	0
Laryngeal Paralysis	0	4	33	0	0	0
Phycogene Dysphonie	0	2	0	41	8	0
Normal	0	0	2	2	45	0
VocalCord Cancer	0	0	0	0	0	13

Performance	Cordectomy	Laryngitis	Laryngeal Paralysis	Phycogene Dysphonie	Normal	Vocal CordCancer
MSE	0.0219	0.0141	0.0253	0.0435	0.04	0.021
NMSE	0.2181	0.1261	0.1715	0.2526	0.23	0.284
MAE	0.1024	0.0934	0.1175	0.1371	0.13	0.108
Min-Abs Error	0.0001	0.0012	0.0026	0.0008	0.00	0.000
Max-Abs Error	0.6443	0.4724	0.7453	0.8622	0.80	0.529
R	0.8883	0.9372	0.9142	0.8661	0.87	0.850
Percent Correct	100	100	94.28	95.34	84.9	81.25

VIII. RESULTS & DISCUSSION

In this research work speech disorder disease were classified using various neural network and SVM. After extracting the relevant features, trainning was done on 75% of samples for more learning and one time testing was done on 25% samples.The epoch applied are shown in table II.The classification was optimum with SVM.The table II shown above gives percentage classification of various diseases using used neural network and SVM.The accuracy for MLP,GFF,Modular network was found out to be 60.30,65.97,52.97 respectively which is not promising except that with SVM which is 92.26%.Table III shows the confusion matrix , performance table,Mean square error, for highest classification in this experiment using SVM .So such accuracy of 92.26 % will help Physicians and Surgeons' to diagnose the considered disease.

REFERENCE

[1] Manjot Kaur Gill,Vector," Quantization based Speaker identification", International Journal of Computer Application(0975-8887)Volume4-No.2, Pp1-4, July 2010.

[2] J.W.Picone,"Signal modeling techniques in speech recognition",Proceedings of the IEEE, Vol.81, No.9, Sept.1993.

[3] LotfiSalhi,TalbiMourad,andAdneneCherif,"Voice Disorders Identification Using Multilayer Neural Network", The International Arab Journal of Information Technology Volume 7-No.2, Pp.177-185 April 2010.

[4] AntanasLipeika, Joana Lipeikiene, Laimutis Telksnys, "Development of Isolated speech Recognition System", INFORMATICA, 2002, Vol 13, No.1, 37-46, 2002.

[5] Milan Sigmund, "Voice Recognition By Computer", TectumVerlag publication, pp no20-22

[6] John G.Proakis, Dimitris, G.Manolakis, "Digital Signal Processing. principles, Algorithm and Applications", Prentice Hall India, pp.309,122 Third Eition.

[7] Orzechowski, A Izworski, R. Tadeusiewiez K. Chmunzynska, P. Radkowski, I Gotkowska, " Processing of pathological change in speech caused by Dysarthria ", IEEE Proceedings of 2005 International symposium on intelligent signal processing & communication system, pp 49-52, Dec 13-16, 2005.

[8] Martinez Cesar E, Refiner Hugo L "Acoustic Analysis of speech for detection of Laryngeal pathologies", IEEE proceeding of the 22nd Annual EMBS international conference chicago IL, pp.2369-2372, 2000.

[9] M.Hariharan,M.P.Paulraj,Sazali, aacob, "Time Domain Features and Probabilistic Neural Network For the Detection Of Vocal Fold Pathology", Malaysian journal Of Computer Science, Vol(23) 2010 pp. 60-67.

[10] Marcelo de Oliveira Rosa,Jose Carlos Pereira and Marcos Gellet,"Adaptive Estimation of Residue Signal for Voice Pathology Diagnose" IEEE Transaction on Biomedical Engineering , Vol.47, No.1, Jan.2000.

[11] Lawrence R. Rabiner,"On the use of Auto correlation Analysis for Pitch Detection",IEEE Transaction Acoustics,Speech and signal Processing",Vol.ASSP-25,No.1,pp24-30,February 1977.

[12] R.W.Schafer and L.R.Rabiner, "System for automatic formant analysis of voiced speech," J.Amer.,vol.47,pp 634-648,Feb.1970

[13] http://en.wikipedia.org/wiki/cepstrum

[14] M.Fernandes, F.E.R.Mattioli, E.A.LamounierJr. and A.O.Andrade,"Assesment of Laryngeal Disorders Through TheGlobal Energy of Speech," IEEE Latin American Transactions, vol.9, No.7, December 2011.

[15] http://www.phon.ucl.ac.uk/resource/sfs/rtgram/AboutSpectrography

[16] SasanAhmadi, Andreas S.Spanians, "Cepstrum-Based PitchDetection using a new Statistical V/UV Classification Algorithm" IEEE Transaction On Speech and Audio Processing ",Vol.7,No.3,pp335-337,May 1999.

[17] Tetsuya Shimamura, Hajime Kobayashi, "Weighted Autocorrelation for Pitch Extraction of Noisy Speech", "IEEE Transaction On Speech and Audio Processing",Vol.9,No.7,October 2001.

[18] ClimentNadeu, Jordi Pascual and Javier Hernando, "Pitch Determination Using The Cepstrum Of The One-Sided Autocorrelation Sequence IEEE Proceedings", pp 3677-3680, 1991.

[19] Manfred Putzer& Jacques Koreman ''A german database for a pattern for vacal fold vibration ''

Phonus 3, Institute of Phonetics, University of the Saarland, 1997, 143-153.

[20] Snell .R.”format location from LPC analysis data “,IEEE transaction on speech an audio processing, 1(2), PP. 129-134,1993.

[21] www.practicalcryptography.com

[22] Coifman, R.R,M.V.Wickerhauser, “Entropy based Algorithm for best basis selection”, IEEE transaction on Information Theory, pp.713-718, 1992.

[23] E.Scheirer& M Slaney.”Construction& evaluation of robust multy feature speech/music discriminator ” in proc.ICASSP,1997.

[24] G.Tzanetakis&P.Cook,”Music genre classification of audio signal “.IEEE Transaction on Speech and Audio Processing ,vol.10,No.5,July 2002.

[25] G.VenBismarck, Timbre of steady sound:A factorial investigation of its verbal attributes. ACUSTICA,30,1974.