Detection of Glottal Pathologic Voice from Speech Signal

Ashish Kumar[1], Ajinkya Biyani[2], Abhay Srivastava[3], Ashish Tomar[4], Prof. S. N. Shelke[5]

[1234](Students, Computer Engineering Department, SIT College of Engineering Kondhwa, Pune)

[5](Professor, Computer Engineering Department, SIT College of Engineering Kondhwa, Pune)

Abstract: **The aim of this paper is to build a better classification models for detection of voice pathologies. Such detection of voice pathology enables objective assessment and earlier intervention for the diagnosis.**

Here we analyse and differentiate the pathological voice from normal voice using data mining technique like Support Vector Machine (SVM). We conducted cross-validation experiments on The Saarbruecken Voice Database using support vector machines (SVM) for classification of normal and pathological voices. The speech signal is analysed to extract the acoustic parameters such as 12 Mel-Frequency Filter Bank Cepstral Coefficients (MFCC) and zero-crossing rate (ZCR). The system gives promising accuracy in the detection of Glottal Pathology.

***Keywords*—Glottal Pathology Detection, Support Vector Machine (SVM), Pathological Voice, 12 Mel-Frequency Filter Bank Cepstral Coefficients (MFCC) and zero-crossing rate (ZCR), Saarbruecken Voice Database.**

I. INTRODUCTION

The malfunction of the human speech production system or auditory system pose a great threat to proper understanding between individuals. Any pathology that occurs to alter the periodic movements of the vocal folds affect speech produced. The previous methods like direct inspection of the vocal folds and the observations of the vocal folds by endoscopic instruments are expensive, risky, time consuming, discomfort to the patients and require costly resources, such as special light sources, endoscopic instruments and specialized video-camera equipment.

Glottal pathology can be detected with speech processing technology. Pathology voice classification model detects the pathologically defective voice signals accurately. The implementation of statistical methods through the ensemble learning concept is used to identify and predict the normal and abnormal voices precisely.
Such detection uses important features of speech signals like Mel-Frequency Cepstral Coefficients (MFCC) [1][2][3][4], Zero-Crossing Rate[2][3], Jitter[4], Shimmer, Pulse, Pitch and entropy [1][3].
The proposed system can help to detect glottal pathologies from the voice input. The system uses MFCC which are vocal tract parameters, in detection of the glottal pathologies and ZCR for the detection. The SVM classifier is used to train the data. The classification of pathologies is carried out with the help of SVM classifier.

II. RELATED WORK

Many researchers have contributed in areas related to the detection of pathological voices.

A. Vocal Features for Glottal Pathology Detection using BPNN

Ashwini Visave et al.[4] uses the discriminative characteristics of speech signal like, pitch, jitter, linear prediction residual and cepstral source excitation to aid such an identification system. Back-propagation Neural Network model is used to classify the glottal pathologic voice from normal voice.

B. Effective Glottal Instant Detection and Electroglottographic Parameter Extraction for Automated Voice Pathology Assessment

Pranav S et al.[5] present an adaptive variational mode decomposition (aVMD) based algorithm for reliable detection of glottal instants and EGG parameters from an EGG signal composed of voiced and non-voice segments. First the determination of the of glottal closure and opening instants from candidates a EGG feature signal is done. The candidate glottal instants are determined by detecting the positive and negative zero crossings in normalized candidate EGG feature signal, respectively. Finally, an autocorrelation features based post processing algorithm is presented to reject non-glottal instants from the non-speech production segments.

C. Cancer larynx detection using glottal flow parameters and statistical tools

Anis Ben Aicha et al.[6] implements a system for for detecting and classifying larynx cancer by investigating different glottal flow parameters. The glottal flow is extracted, temporal and frequency parameters are calculated. From the large obtained set of parameters, the most relevant features are selected for pathologic/normal discrimination. A deep analysis of statistical tools such as boxplot and probability density permits is used to select and ordered the most significant parameters. The detection and the classification of the larynx cancer is achieved by artificial neural network (ANN).

D. Classification of Normal and Pathological Voice Using SVM and RBFNN

V. Sellam, J. Jagadeesan [7] provides a classification of pathological voice from normal voice using Support Vector Machine (SVM) and Radial Basis Functional Neural Network (RBFNN) with the dataset of Tamil phrases. The voice features like Signal Energy, pitch, formant frequencies, Mean Square Residual signal, Reflection coefficients, Jitter and Shimmer are taken into consideration to detect voice disorders in children.

E. Glottal Opening Instant Detection From Speech Signal

Maria Markaki et al. [8] give a joint acoustic and modulation frequency representation, i.e. Modulation Spectrum, of sustained vowel /AH/ for detection and discrimination of voice disorders. The database of sustained vowel recordings from healthy and pathological voices is used, with support vector machines (SVM) for classification having classification accuracy of 94.1%.

III. PROPOSED SYSTEM

We have implementing a system that can analyse and to differentiate pathological voice from normal voice using data mining technique. The detection of vocal fold pathology at an early stage is done from set of features like MFCC and ZCR that are vocal tract parameters. Below figure shows the architecture of the proposed system followed by the system working.

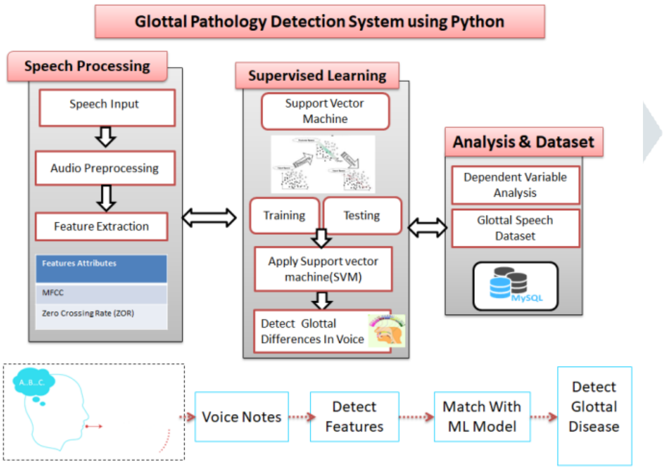

Figure 1: System Architecture

1. **Speech Input:** The main aim of the proposed system is the extraction of glottal parameters from speech signal for the distinction between glottal pathological voice and normal voice. Firstly input speech data is given to the system which contain database of patients suffering from glottal and supra-glottal cancer and from normal persons.

2. **Pre-processing :**The speech data contain lot of noise. By using noise removal techniques or removing any other disturbances present in the data, it is preprocessed to get the fine-tuned data.

3. **Feature Extraction :**

The system can detect vocal fold pathology at an early stage from set of features like MFCC and ZCR, which are vocal tract parameters, in detection of the glottal Pathologies.

a. MFCC (mel-frequency cepstrum (MFC))

The first step in any automatic speech recognition system is to extract features i.e. identify the components of the audio signal that are good for identifying the linguistic content and discarding all the other stuff which carries information like background noise, emotion.

The Mel-Frequency Cepstral Coefficients (MFCC) features is the most commonly used features in speaker recognition. It combines the advantages of the cepstrum analysis with a perceptual frequency scale based on critical bands.

MFCC is based on Human hearing perceptions which cannot perceive frequencies over 1 Khz. In other words, in MFCC is based on known variation of the human ear's critical bandwidth with frequency.

b. Zero Cross Rate(ZCR):

The zero-crossing rate is the rate of sign-changes along a signal, i.e., the rate at which the signal changes from positive to zero to negative or from negative to zero to positive.

The zero-crossing rate is strongly correlated with the spectral centroid, which can be computed using the MFCC spectrum, and is a measure for the high frequency content of a signal.

4. **Classification:** Once the features extracted and normalized, we train a Support Vector Machine model. According to preliminary tests, the best results were achieved. SVM classifier is used for various feature combinations to classify the glottal pathologic voice from normal voice.

IV. ALGORITHM USED

In our Case we are using Support Vector Machine (SVM) to train the model first. We are using Saarbruecken Voice Database for detection. By testing the samples the best results were achieved. SVM classifier is used for various feature combinations to classify the glottal pathologic voice from normal voice.

The reason for selection of this particular algorithm is, it is a powerful classifier that is able to distinguish two classes. SVM classifies the test image in to the class with highest distance up to the neighbouring point in the training.

SVM training algorithm built a model that predict whether the test image fall into this class or another.

SVM necessitate a vast training data to decide a decision boundary and computing cost is very high although we are using single pose (frontal) detection.

The SVM is a learning algorithm for classification which attempt to discover the finest distinguishing hyper plane which minimize the error for unseen patterns.

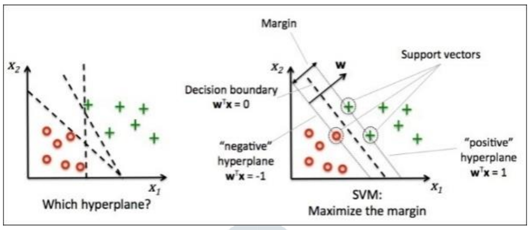

Figure 2: Distinguishing Hyper Plane to Minimize the Error

The data which cannot be distinguished the input is mapped to high-dimensional attribute space where they can be separated by a hyper plane. This projection is well performed by means of kernels.

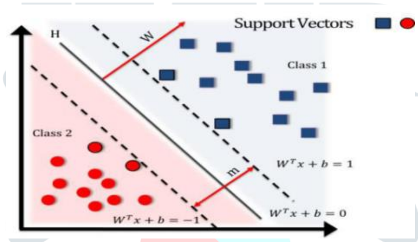

Figure 3: Separating Hyper Plane by Equation

If training set of samples and the equivalent resultant values {-1, 1}. So SVM intend to get the best separating hyper plane specified by the equation WTx+b that make use of the distance between the two classes as shown in above figure.

V. RESULT DISCUSSION

In our proposed system we used a large Saarbruecken Voice Database. Where accuracy and precision are calculated based on false positives images, i.e. which are items incorrectly labelled as belonging to the class and false negatives, which are items which were not labelled as belonging to the positive class but should have been.

1. TP: Positive samples classified as positive.
2. TN: Negative samples classified as negative.
3. FP: Negative samples classified as positive.
4. FN: Positive samples classified as negative

For the mentioned classes the accuracy and precision is calculated by using the formula. The precision is the percentage of documents that are correctly classified as positive out of all the documents that are classified as positive. Where, TP, FP, and FN are truly positive, false positive and false negative images.

$$Precision = \frac{TP}{TP \times FP}$$

$$Recall = \frac{TP}{TP \times FN}$$

Table I: Confusion Matrix

	Class a	Class b	Class c	
Class a	27	1	0	96.42%
Class b	4	24	0	85.71%
Class c	1	0	27	96.42%

The precision obtained by using above for the classes mentioned above is given in below table.

Table II: Precision

Class Label	Precision
a	0.84
b	0.96
c	1

The plot for above table is given in figure 4.

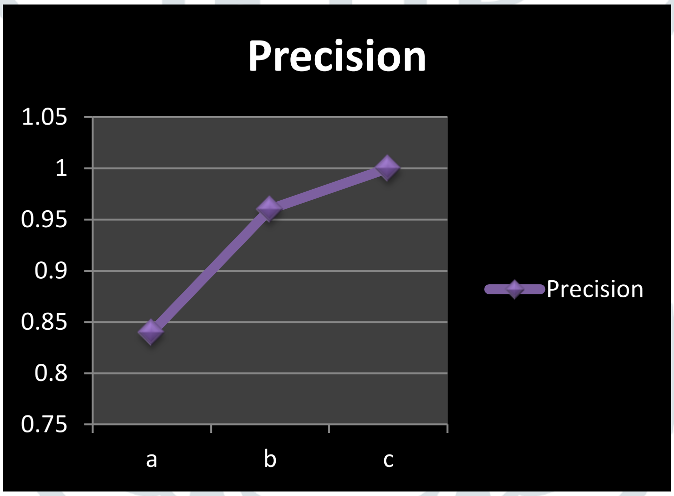

Figure 4: Plot for Precision

Table III: Recall

Class Label	Recall
a	0.96
b	0.85
c	0.96

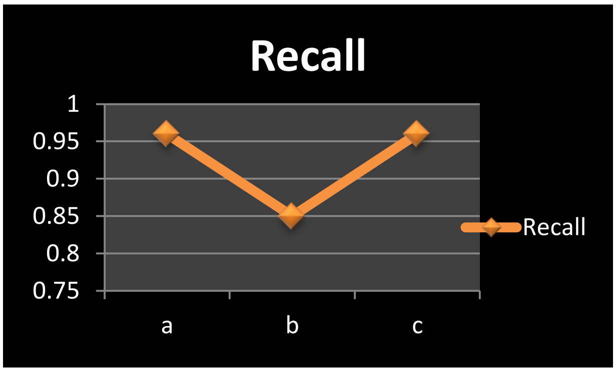

Figure 5.2: Plot for Recall

The accuracy of the proposed system is calculated by using the formula mentioned below.

$$AC = \frac{TP + TN}{TP + FP + TN + FN} * 100$$

The overall accuracy obtained with the help of the confusion matrix is 92.8571 % , where out of 84 i.e. total number of instances the correctly classified instances are 78 and incorrectly classified are 6. We get the 92% accuracy by using SVM with features like MFCCE and ZCR.

VI. CONCLUSION

Speech is one of the fundamental mediums of communicating known to mankind. The absence of such a media will cause a great risk of understanding between individuals. Such a difficulty may arise due to the malfunction of the human speech production system or auditory system.

We are implementing a system that can give a low cost and quick solution by analysing voice source features like 12 Mel-Frequency Filter Bank Cepstral Coefficients (MFCC) and zero-crossing rate (ZCR) in speech data for detection of glottal pathology. SVM classifier is developed for various feature combinations to classify the glottal pathologic voice from normal voice.

VII. REFERENCES

[1] Mythili J, Vijaya MS,"Pathology Voice Detection andClassification Using Ensemble Learning",International Journal of Engineering Science Invention (IJESI),2018.

[2] Pranav S. Deshpande ; M. Sabarimalai Manikandan ,"Effective Glottal Instant Detection and Electroglottographic Parameter Extraction for Automated Voice Pathology Assessment", IEEE Journal of Biomedical and Health Informatics (Volume: 22 , Issue: 2 , March 2018.

[3] S.C.Punitha,N.A. Sheela Selvakumari, "Voice Pathology Identification System using SVM Classifier" ,International Journal of Advance Research inComputer Science and Management Studies,2017.

[4] D. Pravena, S. Dhivya and A. Durga Devi, "Pathological Voice Recognition for Vocal Fold Didease," International Journal of Computer Applications (0975-888), Vol. 47, No. 13, 2012.

[5] N.A. Sheela Selvakumari,Dr. V.Radha, "Recent Survey on Feature Extraction Methods for Voice Pathology and Voice Disorde",International Journal of Computer & Mathematical Sciences,August 2017.

[6] Anis Ben Aicha,"Noninvasive Detection of Potentially Precancerous Lesions of Vocal Fold Based on Glottal Wave Signal and SVM Approaches",International Conference on Knowledge Based and Intelligent Information and EngineeringSystems, KES2018, 3-5 September 2018, Belgrade, Serbia.

[7] Vahid Majidnezhad, Igor Kheidorov, "An ANN-based Method for Detecting Vocal Fold Pathology", International Journal of Computer Applications (0975 –8887) Volume 62–No.7, January 2013.

[8] V.Srinivasan, V.Ramalingam, V.Sellam, "Classification of Normal and Pathological Voice using GA and SVM", International Journal of Computer Applications (0975 –8887) Volume 60–No.3, December 2012.

[9] M.Farra, J. Hernando and P. Ejarque, "Jitter and Shimmer Measurements for Speaker Recognition," TALP Research Center, Department of Signal Theory and Communications, Universitat Politecnica de Catalunya, Barcelona.

[10] Maria Markaki,Student Member, IEEE,and Yannis Stylianou,Member, IEEE,"Voice Pathology Detection and Discriminationbased on Modulation Spectral Features",october, 2010. Aicha Bouzid and Noureddine Ellouze,"GLOTTAL OPENING INSTANT DE- TECTION FROM SPEECH SIGNAL".

[11] V. Sellam, J. Jagadeesan,"Classification of Normal and Pathological Voice Using SVM and RBFNN" ,Journal of Signal and Information Processing, 2014, 5, 1-7, 2014.

[12] Korutla Sudhir Sai , Polasi Phani Kumar," Glottal Analysis Using Speech sig- nals" International Journal for Research in Applied Science & Engineering Tech- nology (IJRASET),2017.

[13] CHRISTOPHE DALESSANDRO and NICOLAS STURMEL, "Glottal closure instant and voice source analysis usingtime-scale lines of maximum amplitude",Sadhan VOL. 36, Part 5, October 2011, pp. 601–622.c, Indian Academy of Sciences.

[14] L Gavidia-Ceballos and J H Hansen,"Direct speech feature estimation using an iterative EM algorithm for vocal fold pathology detection". IEEE transactions on bio-medical engineering, 43(4):373–83, April 1996.

[15] Hemant A. Patil,Pallavi N. Baljekar,"Classification of Normal and Pathological Voices using TEO Phase and Mel Cepstral Features",2012 IEEE.

[16] Ashish Gajanan Nanoti, Jagannath Nirmal,"Glottal Pathology Detection with Vocal Features for Developing Laryngeal Cancer ",Screening Tool Using Neural NetworkConference: International Symposium on Acoustics for Engg. Applications
Organised by Acoustical Society of India (ASI) and KIIT College of Engineering, Gurgaon, At Gurgaon, Volume: NSA031" November 2016.