Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 9, Issue. 5, May 2020, pg.119 – 124

Voice Pathology Classification System Using Machine Learning

Arpitha M S[1]; Dr. Nagarathna[2]
[1]Mtech Student, Department of Computer Science & Engineering, PES College of Engineering, Mandya, India
[2]Professor, Department of Computer Science & Engineering, PES College of Engineering, Mandya, India
E-mail: arpithabharadhwaj@gmail.com, nagarathna@pesce.ac.in

Abstract— Vocal disorders are pathological states that discomfort the quality of speech which is produced by the voice box or larynx. The disrupted nature of voice causes inflammation to larynx or voice box mainly due to overuse or irritation or infection. The goal is to build a machine learning model which categorizes distinct class of diseased condition of vocal chords, include Dysphonia or spasmodic dysphonia, Normal, Stammering which is an instance of stuttering and Vocal palsy or vocal fold paralysis from AIISH (All India Institute of Speech and Hearing), voice data repository and voice from individuals. The machine learning classifiers used to handle the classification problem of vocal disorders are Support Vector Machine and K-Nearest Neighbor. The resulted outcome is evaluated based on the features of voice selected from the process of feature extraction using MFCC (Mel Frequency cepstral coefficients).

Keywords— Vocal disorders, Voice, Machine Learning, AIISH, Feature Extraction, MFCC

I. INTRODUCTION

Vocal disorder is a problem when there is abnormality in the vibration of vocal cords. Voice problem is developed for many reasons. The most common causes for vocal disorders are screaming, smoking, stress, inflammation and swelling, extra growth in vocal cords, alcoholic addiction, nerve problems, hormones, misuse of voice etc., may affect the normal speech ability which includes hoarseness, strained sound, pain or lump while speaking [1,2]. The common vocal disorders are dysphonia, vocal fold paresis or vocal palsy, stuttering or stammering and so on. Dysphonia is a clinical condition where the voice hears to be shivered and creeky, it is mainly due to neurological defect which affects the ability to speak. The very common symptoms of dysphonia are dry cough, impaired voice and sneeze. It can be cured by injecting botulinum toxin [3]. Vocal palsy is a severe voice disorder occurs when voice box is disrupted due to paralysis because of inability of one or both the vocal folds to move, it even affect breathe in addition with ability to speak. Sometimes this condition is also called as vocal folds, this is caused when nerves are damaged due to nerve impulses in the vocal cord. There is no home treatment for vocal cord paralysis and this should be treated under the guidance of medical professionals. The tools used to examine vocal cords may

include, acoustic analysis, laryngeal electromyography, mirror, flexible, rigid and video laryngoscope. Stammering is also known as stuttering is a very common vocal disorder occurs in all age groups and it is temporary in children at the age between 2 and 5. This is recognized by repetitions or stretches of repeated sounds or words. Common causes for stuttering include family history, nervousness, emotional and stress. Stammering in adults happens as a result of brain injury or stroke. It can also caused by emotional trauma [4]. Many researches has been carried out for the classification of voice diseases from past so many decades. The movement of vocal folds can be visualized by a specialized technique called phonovibrography. For the diagnosis of vocal disorders PVG features are extracted by considering the high speed recordings of vocal fold by assessing videoendoscopy [5]. For the diagnosis of pathological voice, Hidden Markov Model and Gaussian Markov Model were used to distinguish the mixed category of diseased and normal characteristics of voice by considering the parameters such as NHR, SPI and RAP, the malfunctioning of vocal cords can be detected.

II. LITERATURE SURVEY

According to previous studies [6] pathological affected voices are detected by using Ensemble learning techniques, it consists of wavelet packet decomposition and linear predictive coding. Disorder detection was carried out by using Bagging algorithm, in which inputs were generated by bootstrapping technique and F-measures were compared for classification of disorder affected samples. The hoarseness in voice changes from children to adults and the causes for hoarseness varies as age group changes since the size of vocal nodule varies as age differs [7]. The significance of hoarseness leads to vocal cord paralysis and it is diagnosed by micro laryngeal examination. Alcoholism is the main risk factor for hoarseness in voice. Parkinson's disease occurs mainly due to defect in the central nervous system. It affects speech in addition with passive movements of the body. Vowels, back vowels and numbers are considered to identify Parkinson's disease. To classify between normal people and Parkinson's people feed forward algorithm is implemented which consists of feed layers such as input, hidden and output [8]. Detection of multiple voice pathological states is investigated by using acoustic voice analysis. To accomplish this work, a system is developed with a supervised algorithm. The database was collected from phonetics and ENT clinic. The entire work is simulated using MATLAB for training and testing [9]. Speaker emotions are taken to identify the state of patient's disease. Based on pattern recognition, classification is done. Multiple speech disorders are very common nowadays. The people diagnosed with dysphonia may have different laryngeal pathologies (LP). Spectral features are used in this model to formulate voice spectrum [10]. The abnormal behaviors observed are glottal stops, wakening of breathiness and screaming with harshness.

III.PROPOSED METHOD

Machine learning system is developed to classify the vocal disorders where the real time voice samples are the input for the system, which is pre-processed to remove noise from the original sample [2]. After removal of noise, the system is subjected to feature extraction to extract necessary features from the sample and selected features are passed through machine learning classifiers to classify the vocal disorders as shown in Figure 1.

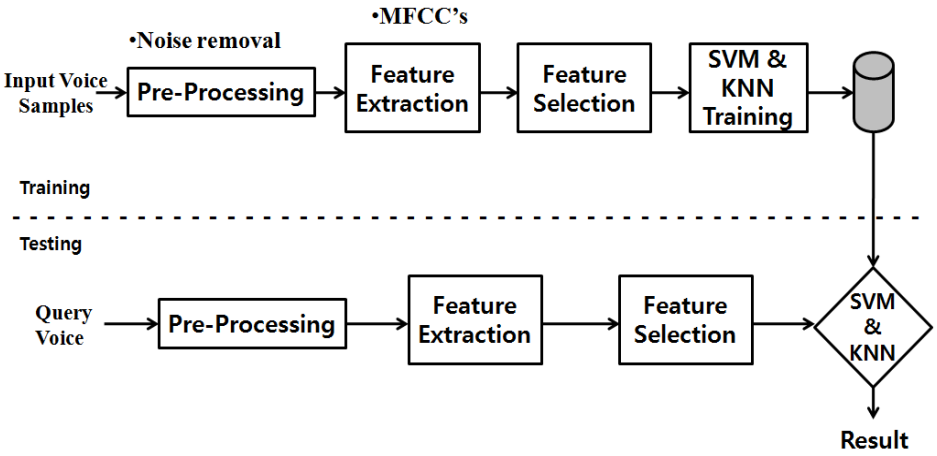

Figure 1: The workflow for classification of vocal disorders

Real time **Voice Samples** are taken as input for the system, which are collected from AIISH (All India Institute of Speech and Hearing). The recorded voice samples are of four categories namely, normal (healthy voice), dysphonia (shivered), stammering (stretched voice) and vocal palsy (paralysis to vocal chord). The minimum length of each sample is trimmed to 5 seconds and saved in .wav format, the wave representation of voice sample is shown in Figure 2.

Figure 2: The waveform of input signal of normal voice

Audio Pre-Processing is the primary process to remove noise from the original voice samples. The filter used for this work is a non linear Median filter, since there exists lot of variations in voice samples of different categories. Filter eliminates unwanted sections of voice signal to enhance its quality. Audio file is made read using audioread(a), where 'a' is a required audio clip. The input data is passed through Median filter, medfilt(SampleData, SampleRate) where SampleData is the number of samples per second and SampleRate is the number of bits per sample. Filter design is made in order to carryout mirroring operation that consists of a) filter order by calculating maximum delay b) zero pole gain conversion and c) second order stability for digital filtering. Mirroring function, filifilt(soslp, glp, signal) performs filtering by processing the input data in both forward and reverse direction. Smoothed signals are obtained by modifying the data points of a signal, plot(filtered_sound) plots the resultant filtered smoothed signal as shown in Figure 3.

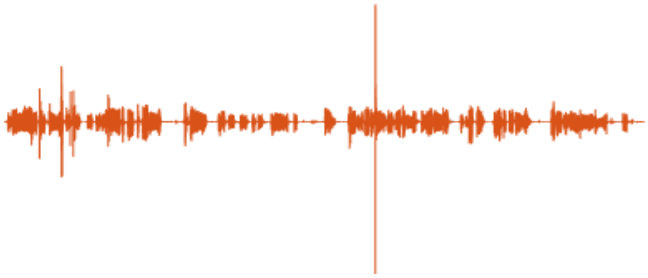

Figure 3: The waveform of filtered signal of normal voice

In the process of **Feature Extraction**, the features of training set of signals are extracted and compared with testing sample set. To carry out this process, loading of data is the initial step and calculation of mel-frequency cepstral coefficients with frequency 'fs' for training set and testing set are performed separately and store in ceptral cell. To calculate mel ceptrum of a signal 'c', sensible combination of window should be selected, which includes, rectangular window in time domain 'R', Hanning window 'N', and Hamming window 'M'. The selection of window depends on the nature of the signal. After selecting window, the function 'ENFRAME' splits the signal into overlapping frames in the order of one frame per row. The number of frames is fixed to fix((length(X)-LEN+INC)/INC. Discrete Fourier Transform of real framed data is calculated and data is padded or truncated to length 'N'. Melbank determines matrix for a mel-spaced filterbank which outputs a sparse matrix containing the filterbank amplitudes. Power spectrum of signal is calculated and discrete cosine transform of real data is obtained which is truncated to length 'N'. The selected features are a) minimum 'mn' (minimum value of the entire row of the matrix), b) first quartile 'fst' (rows from one to four of the matrix), c) third quartile 'trd' (rows from five to eight of the matrix), d) median 'md'(median values of the entire row of the matrix) and e) maximum 'mx' of each row of a resultant sparx matrix. The extracted features are shown in Figure 4.

	Min	First	Third	Median	M
1	-1.4407	1.5878	-0.3859	0.0531	
2	-4.6838	1.3687	0.0313	0.7645	
3	-1.7680	0.4987	-0.1946	0.3565	
4	-2.2448	2.2078	-0.5036	0.2255	
5	-2.4713	3.7953	0.0421	0.0502	
6	-2.0898	1.8762	-0.1782	-0.0789	
7	-3.1496	3.0636	0.2032	0.2351	

Figure 4: Selected features of feature extraction

In the process of **Classification**, the browsed (input) voice sample is classified into respective voice disorders based on the features of that particular voice. To carry out this, Multi Support Vector Machine (mSVM) is used, which does the operation by considering one to one object feature. The parameters for mSVM are training matrix 'T', testing matrix't' and group 'C'. The common kernel functions that can be chosen for SVM are linear, polynomial and radial basis. In this work, since the classification is non-linear, radial basis function (rbf) is selected and it is defined as $k(x , x') = \exp(-(\|x-x'\|^2) / 2\square^2)$ where $(x - x')$ is the Euclidian distance between two data points x and x' respectively. The gamma value is set to 'high' to make decision boundary thicker (curvy). To improve the accuracy of the system, one more

classifier is used i.e, K-Nearest Neighbor (KNN). This is an instance based algorithm which assumes all instances to points. The nearest neighbor is defined in terms standard Euclidean distance as $<a_1(x), a_2(x),\ldots a_3(x)>$ where $a_n(x)$ is the value of n^{th} attribute of x instances. The distance between the instances x_i and x_j can be defined as $d(x_i , x_j)$, where $d(x_i , x_j) = \sqrt{\sum_{r=1}^{n}\left(a_r(x_i) - a_r(x_j)\right)^2}$.

IV. RESULTS

The result in Table I shows the classification of vocal disorders such as normal, dysphonia, stammering and vocal palsy. The disorders are differentiated based on the values obtained during classification. The average values of the classes are 10.5116, 7.56655, 8.81489 and 2.67294 for normal, dysphonia, stammering and vocal palsy respectively.

TABLE I

Class	Value
Normal	10.5116
Dysphonia	7.56655
Stammering	8.81489
Vocal Palsy	2.67294

The result in Table 2 shows that k-nearest neighbor (knn) has good accuracy rate and correct classification rate as compared to that of multi support vector machine (mSVM).

TABLE II

Class	Total Samples	SVM		KNN		Accuracy	
		True	False	True	False	SVM	KNN
Normal	40	35	5	38	2	0.874	0.950
Dysphonia	40	22	18	30	10	0.550	0.750
Stammering	47	39	8	44	3	0.829	0.936
Vocal Palsy	38	31	7	35	3	0.816	0.921

The true count and false count of each label (class) is displayed in the above table 2. This shows that normal has highest true count on knn and dysphonia has lowest true count on svm.

V. CONCLUSION

In this study, it is observed that all the classes except dysphonia are found decent. The result from Table 2 shows that Dysphonia is difficult to detect for svm algorithm, where it has accuracy below 60% and it is quite improved on knn and reached accuracy of 75%. To further improve the quality of system, dysphonia might be a good option. Here, in this work, four classes are chosen for classification by taking average of 40 samples each, it can be enhanced further by choosing large number of samples and classes.

REFERENCES

[1] Arati Alva, Megna Machado, et al, and Suja Sreedharan, "Study of Risk Factors for Development of Voice Disorders", Journal of Clinical and diagnostic research: JCDR 11(1), MC01, 2017.
[2] Arpitha M S, Nagarathna, "A Frame Work for Classification of Vocal Disorders without Clinical Intervension", IJCSE, Vol 8, Issue 1, Jan 2020, p 70-73.
[3] Ugo Cesari, Giuseppe De Pietro, Elio Marciano, Ciro Nirli, "Voice Disorders Detection via an m-Health System", BioMed Research International, Volume 2018.
[4] Paulo Eduardo Przysiezny, Luciana Tironi Sanson Przysiezny, " Work-related voice disorder Disturbio de voz relacionado ao trabalho", Brazilian Journal of Otorhinolaryngology, Vol 81, Issue 2, March-April 2015, p 202-211.
[5] Voigt D, et al. "Classification of functional voice disorders based on phonovibrograms", Artif Intell Med, 2010 May, 49(1) : 51-9.
[6] Mythili J, Vijaya M S, "Pathology Voice Detection and Classification Using Ensemble Learning", IJESI, Vol 7, Issue 8, Aug 2018.
[7] Edakkattil Rameshkumar, Tony Kalliath Rosmi, "Prevalence of age, gender and pathological conditions of vocal cords leading to hoarseness of voice", IJAM, May 2016.
[8] Akshay S, Kiran Vincent, "Identification of Parkinson Disease Patients Classification using Feed Forward Technique based on Speech Signals", IJEAT, Vol-8, Issue-5, June 2019.
[9] Mohamed FEZARI, Fethi AMARA and Ibrahim M.M, "Acoustic Analysis for Detection of Voice Disorders Using Adaptive Features and Classifiers", Proceedings of the 2014 International Conference on Circuits, Systems and Control.
[10] Juan Rafael Orozco-Arroyave, et al, "Characterization Methods for Detection of Multiple Voice Disorders", IEEE Journal of Biomedical and Health Informatics, 2015.

