A Survey Paper on Detection of Voice Pathology Using Machine Learning

Yashaswini D M[1], Dr. Bharathi M[2]

[1]Student, S J C Institute of Technology, chikballapur, Karnataka, India
[2]Professor, S J C Institute of Technology, chikballapur, Karnataka, India

---***---

Abstract - *Voice pathology detection systems can significantly contribute to the assessment of voice disorders as well as provide early detection of voice pathologies. These systems used machine learning techniques, regarded as highly promising in the detection of vocal pathologies. The goal of this research is to create a powerful feature extraction voice pathology detection tool based on Deep Learning.*

This paper describes a system for detecting voice pathology that uses machine learning to categorize voice signals as healthy or pathological. Various features such as Mel-Frequency Cepstral Coefficients (MFCC), Linear Predictive Analysis (LPC), Wavelet Packet Decomposition (WPD), Cepstral Analysis (CA), Jitter, Shimmer, Pulse, Pitch, Harmonicity, Intensity, Energy, and Entropy are extracted during this phase. Three widely used measurements are used to evaluate the proposed method: accuracy, sensitivity, and specificity.

Key Words: Voice pathology, MFCC, CNN, LPC, SVM

1. INTRODUCTION

Machine learning skills are extremely valuable in classifying at least two classes, especially in speech processing. Furthermore, these methods have been employed in a range of healthcare environments. Voice pathology detection is one of these medical applications. If a person has difficulty or hoarseness in their speech as a result of a speech organ defect, psychological disorder, accident, autism, or other conditions, they are considered to have a vocal pathology problem. The existence of voice disorders or abnormalities in the vocal tract interferes with the glottis' normal vibrating sequence, resulted in hoarseness. For detecting vocal pathology, traditional medical diagnostic procedures are useless. These techniques are mainly based on vocal cord examination, which can lead to confusion and incorrect assesments. Furthermore; these methods necessitate a wide range of equipment, take more time, and are inefficient in terms of cost. One of the most difficult research areas in speech analysis and processing is the mechanisms of pathological voice detection and classification.

Voice disorder detection techniques have advanced significantly thanks to the use of machine learning algorithms, which have demonstrated their efficiency and productivity throughout diagnostic applications in recent years, particularly in voice pathology detection systems These algorithms' primary function is to evaluate and recognise the voice disorder, and then to build a system capable of dealing with sound waves pathologies in order to differentiate pathological voices from healthy voices.

2. OBJECTIVE

The primary goal of our proposed method is to develop voice pathology detection systems that can effectively contribute to the assessment of voice disorders and provide early detection of voice pathologies.

3. PROBLEM STATEMENT

Suggested research has demonstrated that anomaly based detection techniques could contribute effectively towards the evaluation of voice disorders and provide earlier diagnosis of voice Pathologies. These systems used machine learning techniques that are thought to be quite promising for detecting voice disorders. However, the majority of proposed systems for detecting voice disorders made use of a limited database. Furthermore, one of the most difficult issues for these techniques is their low accuracy rate.

Several difficulties in speech synthesis computation have now been discussed, which include comprehension evaluation, microphone validation, and identification of para linguistic events in voice, including such emotional responses and pathologies.

4. LITERATURE SURVEY

The authors of [1] did work on vocal signal processing techniques in order to detect voice disorders. The extracted voice samples have been focused on glottal signal parameters, and SVM and K-Nearest Neighbor Boring classifiers were used (k-NN).The SVD database is used for voice samples in this method, with 71 pathological and 34 healthy voice samples collected, respectively. SVM was 98.5 percent accurate, while K-NN was 88.2 percent accuracy, according to the results. However, both the healthy and pathological voice samples are considered small.

In [2] explored and investigated the symptom severity disease voice using CNN for classification and Fourier-based synchro squeezing transform (FSST) for feature extraction.

The FSST is a technology used to increase the size of data. The SVD database was used to obtain the voice recordings for regular and anomalous voices. The database is well-balanced in terms of group size (i.e., 94 voice samples for the normal and abnormal group). The results, however, really aren't encouraging, also with accuracy rate just hardly reaching 70%.

In this regard, [3] proposes a new technique for fully automated recognition and characterization of speech pathology. The properties of peak values, entropy, and lag are analysed and extracted from voice samples in this system using the autocorrelation technique. These characteristics are then fed into an SVM classifier, which detects and categorises speech pathology. The three datasets used were MEEI (101 pathological samples and 53 healthy samples), SVD (263 pathological samples and 266 healthy samples), and AVPD (101 pathological samples and 53 healthy samples) (127 and 169 for healthy and pathological samples, respectively).

The authors picked the SVD voice database for patients with spasmodic dysphonia in [4]. In total, there are 100 voice samples (40 pathological and 60 normal voices). They also used MFCC to extract characteristics from speech signals and combined two types of classifiers, SVM and GMM, to distinguish between healthy and ill voices. They also used Bhattacharyya (Bh) and Kullback-Leibler (KL) distance measurements to measure the GMMs' distances in order to increase their discriminative abilities. According to the data, as compared to traditional Bh and KL distances, this method enhanced performance by 4% and 2%, respectively, in terms of sensitivity. GMM-SVM has a degree of accuracy of 96.6 percent. However, this method has only been tested on a small scale database.

In [5] outlines a framework for healthcare monitoring tracking that is linked to the city of the future and smart gadgets in order to ensure that healthcare is accessible and affordable. The authors describe a Voice Pathology Detection (VPD) approach using electroglottography (EGG) and voice signals as inputs. The input devices are connected to the Internet, and the data they generate is sent to the cloud. These signals are then investigated and classified as aberrant or normal. The results of the signals are sent to the doctors, who will decide what to do next. The database and categorization are based on the Saarbrucken Voice Database and the Gaussian mixture model, respectively (SVD). The proposed method has a higher level of precision.

4. PROPOSED SYSTEM

Audio will be selected from the system or recorded using any audio recorder, then pre-processing will be performed to remove noise from the input audio, feature extraction will be performed using one of several feature extraction methods, and the extracted features will be cascaded and fed to a trained machine learning algorithms for classifying of Voice Pathology Detection.

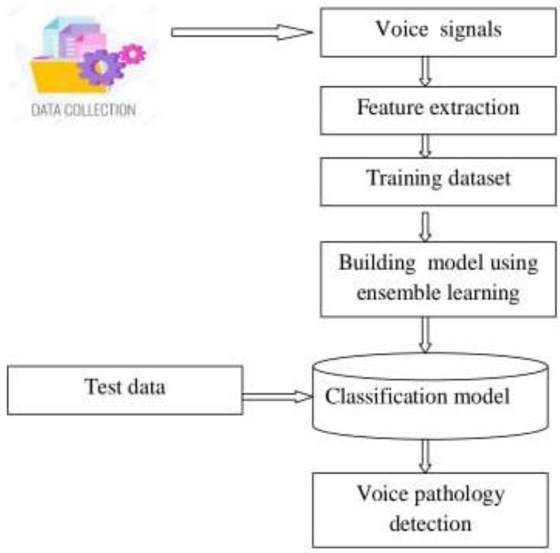

Fig -1: Flow diagram

5. CONCLUSIONS

According to the findings of this study, understanding the pathology in terms of its genesis, as well as the organs or tissues involved in the illness and the speech production process, is required before categorising voice recordings. After these investigations, the characterization will no longer be blind, allowing for a more appropriate selection of measurements. Periodicity characteristics, for example, may be the best option if the disease is directly related to the stability of vocal fold vibration; nevertheless, if indeed the issue is connected to dysphonia, noisy contents characteristics or spectral and cepstral analysis should have been the best option. In any case, the speech therapist may benefit from making an informed decision about the approaches utilised to represent the auditory signals. Helping clinicians make more accurate therapy and/or treatment choices for patients.

REFERENCES

[1] V. Mittal and R. k. Sharma, "Glottal Signal Analysis for Voice Pathology," 2019 2nd International Conference on Innovations in Electronics, Signal Processing and Communication (IESC), Shillong, India, 2019, pp. 54-59, doi: 10.1109/IESPC.2019.8902368M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[2] A. Rueda and S. Krishnan, "Augmenting Dysphonia Voice Using Fourier-based Synchrosqueezing Transform for a CNN Classifier," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing

(ICASSP), Brighton, UnitedKingdom,2019, pp. 6415-6419, doi: 10.1109/ICASSP.2019.8682391.

[3] A. Al-Nasheri, G. Muhammad, M. Alsulaiman, Z. Ali, K. H. Malki, T. A. Mesallam, and M. F. Ibrahim, "Voice Pathology Detection and Classification Using Auto-Correlation and Entropy Features in Different Frequency Regions," IEEE Access, vol. 6, pp. 6961-6974, 2018, doi: 10.1109/ACCESS.2017.2696056.

[4] F. Amara, M. Fezari, and H. Bourouba, "An Improved GMM-SVM System based on Distance Metric for Voice Pathology Detection," Applied Mathematics & Information Sciences, vol. 10, pp. 1061-1070, May 2016.

[5] M. S. Hossain, G. Muhammad, and A. Alamri, "Smart healthcare monitoring: a voice pathology detection paradigm for smart cities," Multimedia Systems, pp. 1-11, Jul. 2017.