International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 05 | May - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Detection of Voice Disorder Using Machine Learning

Tejal Sawdekar[1], Pooja Kasar[2], Mansi More[3] Anuja Rane[4]

[1]Tejal Sawdekar Department of Information Technology & Met Institute of Engineering
[2]Pooja Kasar Department of Information Technology & Met Institute of Engineering
[3]Mansi More Department of Information Technology & Met Institute of Engineering
[4]Anuja Rane Department of Information Technology & Met Institute of Engineering

---------- --***--

Abstract - Voice disorder is a health issue that is frequently encountered, however, many patients either cannot afford to visit a professional doctor or neglect to take good care of their voice . In order to give a patient a preliminary diagnosis without using professional medical devices, previous research has shown that the detection of voice disorders can be carried out by utilizing machine learning and acoustic features extracted from voice recordings. Considering the increasing popularity of machine learning and feature learning, this study explores the possibilities of using these methods to assign voice recordings into one of the two classes Normal and Pathological . While the results show the general viability of machine learning and for the automatic recognition of voice disorder, they also demonstrate the shortcomings of the existing datasets for this task such as insufficient dataset size and lack of generality. The best accuracy in voice diseases detection is achieved by the(CNN). The proposed work uses spectrogram of voice recordings from a voice database as the input to a Convolutional Neural Network (CNN) for automatic feature extraction and classification of disordered and normal voice.

Key Words: Voice recognition, Defective speech, Convolutional Neural Network, short time fourier transform, vocal fold.

1.INTRODUCTION

Voice is the basic human instinct and voice one among its subsystem. Voice disorder deviates quality, pitch, loudness and vocal flexibility from voices of similar age , gender and social groups. voice disorder includes neoplas, phonotrauma and vocal palsy. Voice Disorders are medical conditions involving abnormal pitch, loudness or quality of the sound produced by the larnyx i.e. voice box. A voice disorder can cause an irregular movement of the vocal folds during phonation. Voice disorders are pathological conditions that directly affect voice production. Most of the traditional diagnostic methods on voice disorders rely on the expensive devices and clinician's experience. These methods result not only in considerable cost for the patients and incorrect detection of diagnosis, but also cause delay for the patients in places without the specialists and medical resources. Computer aided medical systems are being used more and more often to help doctors diagnose the pathological status of the voice with lower cost and non subjectivity (not subjected to a doctor's personal experience) In the sense of data science, the diagnosis of pathological voice is a multiclass classification problem that is dependent on the audio signal from individuals. The diagnosis of pathological voice is a multiclass classification problem that is dependent on the audio signals from individuals. The detection of voice disorders can be carried out by utilizing machine learning and features extracted by voice recordings. The goal of this study is to create machine learning models that can not only detect voice disorder, but also correctly categorize the type of voice disorder from the voice samples provided by the dataset. This paper presents a new system of pathological voice recognition using convolutional neural network (CNN) as the basic architecture. The new system uses spectrograms of normal and abnormal speech recordings as input to the network. CNN effectively extracts features from diagnose voice disorders and also makes the system more robust. In this paper, we use the CNN convolutional neural network structure for automatic analysis of voice recording spectrograms.

2. METHODOLOGY

2.1 Input data

One of the properties of CNN is the ability to reduce the dimension of two-dimensional feature maps. Therefore, speech recordings are converted from one-dimensional signals to two dimensional spectrogram.

2.2 Dataset

This paper uses the Saarbruecken voice database, which was registered by the Institute of phonetics of the Saarland University in Germany. This database contains 71 different pathologies with speech.

2.3Data Preprocessing

To use CNN for application, a 2-Dimensional graph is ideal for extracting features. In this case, we need to perform some pre-processing steps to form the feature map to feed into the CNN system. The original speech is first resampled at 25 kHz .Furthermore, Short-Time Fourier Transform (STFT) are applied to the resampled data for transforming the time domain signal into spectral-domain signal. In STFT, each file use 10 ms hamming window segments, with 50 percent overlap between consecutive windows. Finally, the spectrograms are reshaped to the same size of 60*155 points, with 155 being the minimum length of the spectrograms. This is because there is a large part of the area in the spectrogram contains no information, and the useless part of the spectrogram is cut off to reduce the effect of noise to the classification result.

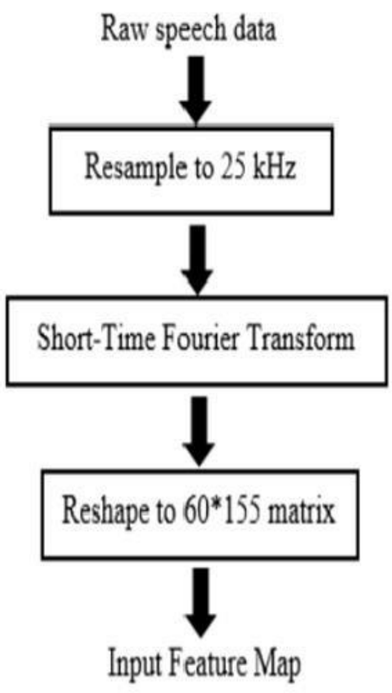


Fig 1. Data Preprocessing

2.4 Feature Extraction

Pathological voice contains subtle differences that can be seen on the spectrogram compared to normal voice, which are difficult to be manually defined using particular criteria. Hence the CNN plays an important role as a feature extractor. One of the properties of CNN is the ability to reduce the dimension of two dimensional feature maps. Therefore, speech recordings are converted from one dimensional signals to two dimensional spectrograms.

2.5 CNN Algorithm

For deep diagnosis of the voice disorder, we will use Convolutional neural network (CNN) as a deep learning method. Convolutional Neural Network is a widely popular deep learning architecture.

A general CNN is comprised of one or more convolutional layers and max pooling layers followed by one or more fully connected layers. CNN have been extensively applied to image classification tasks but also have been increasingly popular for audio analysis such as voice disorder Detection. The size of the input feature map is 60*155*1. Since it is the spectrogram of the speech file, the depth of this input layer is 1. The input feature map is then convolved with a set of 8 filters. Each filter has the shape of 8*3*1 and stride of 1. We use the rectangular filters in this work due to the spectrogram characteristics. Furthermore, max-pooling filters with the shape 4*4 and stride of 1 are applied to pool the significant values out and reduce the computational complexity. Then the activation function RELU is applied to make the neural network non-linear and fit for classification. After the first hidden layer, each layer was convolved with 8 filters with the shape 8*3*8 and stride of 1. Max-pooling filters and activation function is the same as for the first hidden layer. After 10 hidden layers to extract the features from the spectrogram, the feature map is formed into a Dense Layer, which is a fully-connected layer, to train the model for classification.

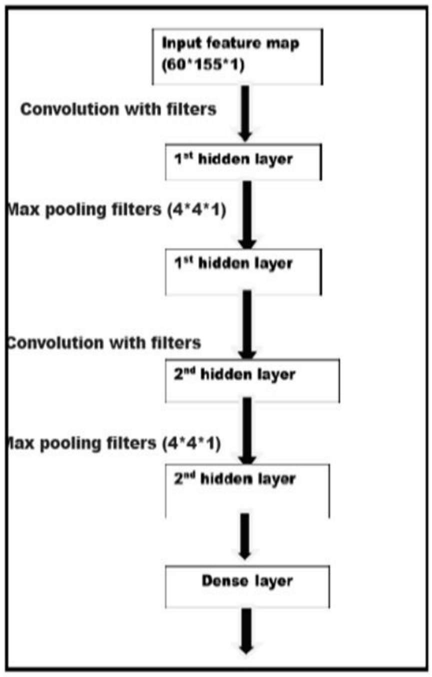


Fig 2. CNN Algorithm

3.TASK FLOW

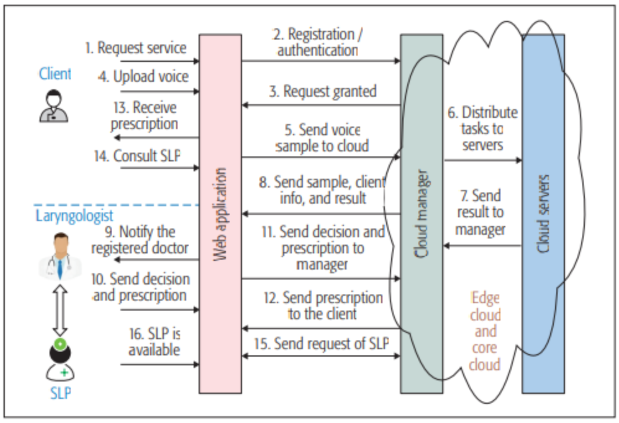
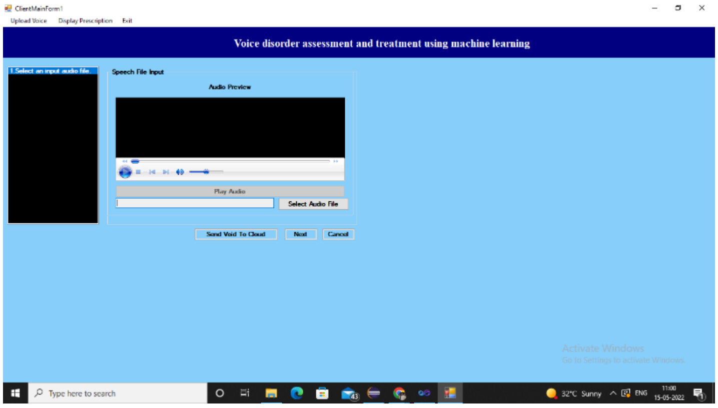


Fig 3. Task flow of the proposed voice disorder detection and treatment

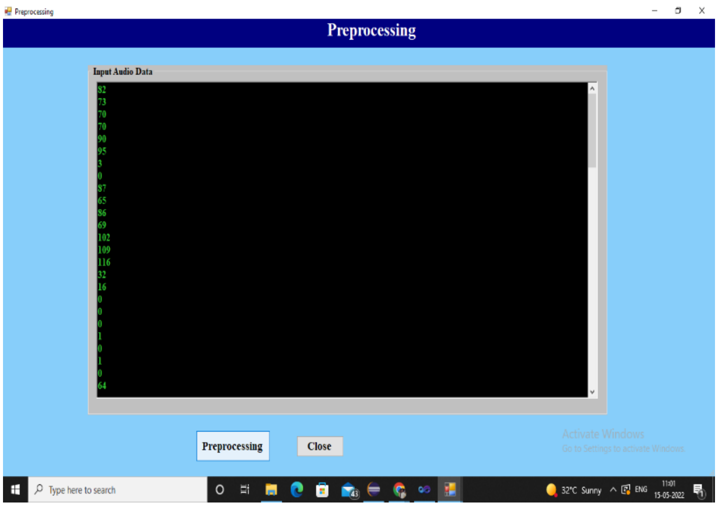
The task flow of the proposed framework is shown in Fig. 3. There are 16 consecutive tasks to complete the whole process. The client first requests a service through a web application. The EC manager verifies the authenticity and grants the request. The client then uploads his or her voice sample to the application. The sample is distributed to different cloud servers, who work in parallel. The voice disorder assessment and treatment system, which involves the servers, sends the result to the manager. The manager sends the sample, result, and client info to the registered specialist for a final decision on the assessment. The doctor prescribes some therapy or medicine, which is sent to the client. If the client needs an SLP, he sends a request through the web application, and the SLP is notified through the cloud manager. The SLP may visit the client for speech therapy or conduct the therapy from a remote location. In the case of conducting the therapy from a distant location, the treatment needs another automated system that can have a communication framework of speech samples without any distortion. The treatment can also be in the form of a humanoid robot, who will be instructed to interact with the patient for therapy or rehabilitation. It has been observed that patients often do not show up to hospitals for post-diagnosis management purposes if they feel better. A humanoid robot can therefore be engaged for follow-up of the patients.

4.RESULT

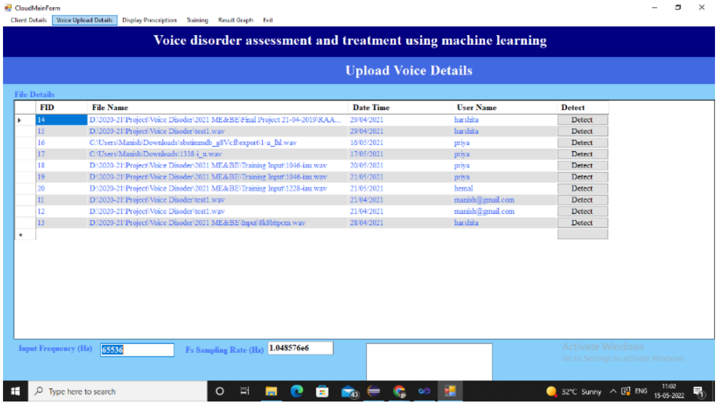
In the below given screenshot after login to client one can upload the voice only in waveform and then send to cloud and in the screenshot shown below the voice is uploaded successfully.



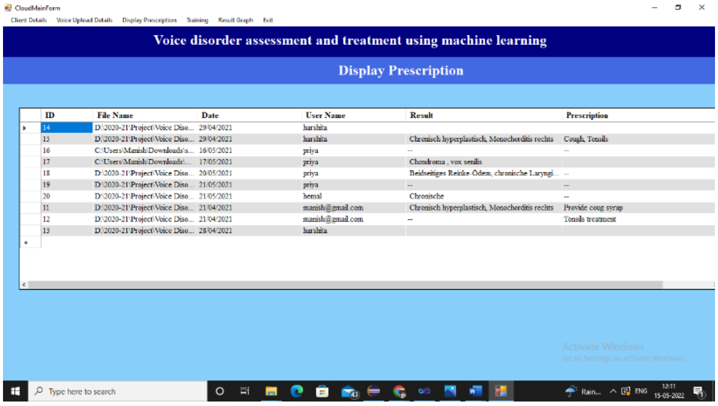
In the below given screenshot Preprocessing after uploading the voice is shown. The data or the number displaying in the box is the bytes of the uploaded voice after dividing into the parts.



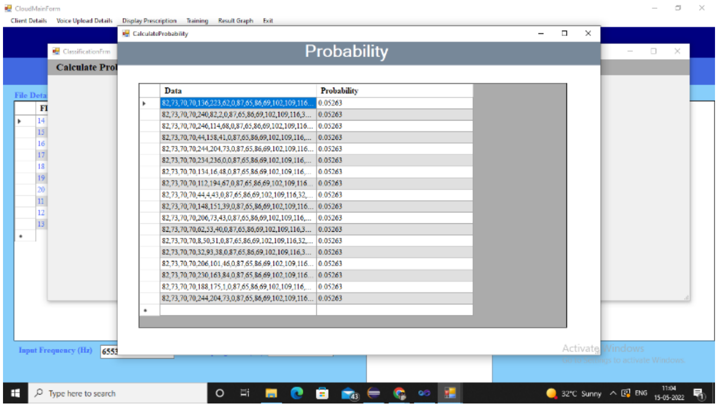
In the below screenshot uploaded voice details is been shown. From here one can also detect and get details about their disease.



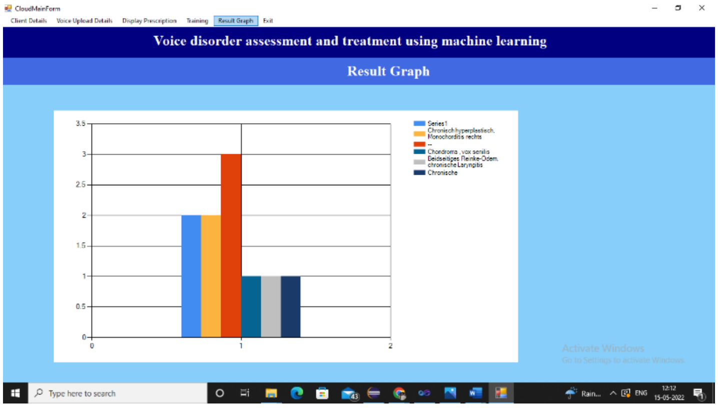
The image displays prescription for the client who have already registered in data cloud and are the part of it and along with that provides the prescription.



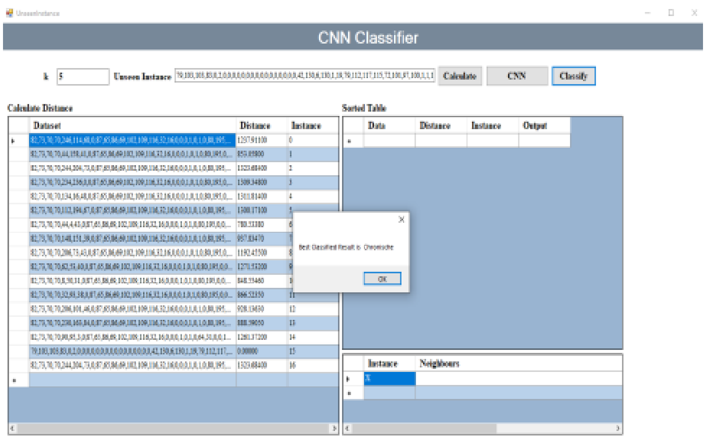
Here the screenshot shows the details about classification that how it works after login into the data cloud and gives the detail information about the probability, in probability whatever voice data you store is converted in some kind of bytes as given in above image



The shown images gives the information about training data and the graph , where in training data one can add voice data and train it into the proper form then sent to the cloud and graph statistically shows the result about clients in order to their respective disease.



The above image shows CNN Classifier, that after classifying the algorithm it gives the best result as shown.



5. CONCLUSIONS

This paper introduces a novel and real time system for voice abnormalities detection using a deep learning Convolutional Neural Network (CNN) model. The model has been implemented in a voice abnormalities detection system. The development methodology of the voice abnormalities detection system comprises of dataset preparation, learning process, training and validation. The main contribution of this work is modelling a deep learning CNN to provide accurate voice abnormalities detection results.

6.REFERENCES

[1] Minh Pham Diagnosing Voice Disorder with Machine Learning Center of Urban Transportation Research,2018.

[2] Shia SE, Jayasree T. Detection of pathological voices using discrete wavelet transform and artificial neural networks. 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS). IEEE; 2017.

[3] Jothilakshmi, S. Automatic system to detect the type of voice pathology[J]. Applied Soft Computing, 2014, 21: 244- 249 .

[4] M. S. Hossain and G. Muhammad, "Healthcare Big Data Voice Pathology Assessment Framework," IEEE Access, vol. 4, no. 1, 2016, pp. 7806–15.

5] G. Muhammad et al., "Voice Pathology Detection Using Interlaced Derivative Patterson Glottal Source Excitation," Biomedical Signal Processing and Control, vol. 31, Jan.
2017.

[6]https://medium.com/x8-the-aicommunity/audioclassification-using-cnncoding-examplef9cbd272269e

[7] Wu, Huiyi, et al. "A deep learning method for pathological voice detection using convolutional deep belief networks." Interspeech 2018 (2018).