IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁘ Impact Factor 7.918 ⁘ Vol. 12, Issue 2, February 2023

DOI: 10.17148/IJARCCE.2023.12224

Pathologies Classification in Voice Signal Using Deep Learning Technique

Prof.Manjunatha T N[1]

Dhanush M R[2], Devendra kumar C A[3], Abhishek M V[4], Harshith R[5]

Assistant Professor, Computer Science, East West Institute of Technology, Banglore, India[1]

Student, Computer Science, East West Institute of Technology, Banglore, India[2-5]

Abstract: Many references support the non-invasive detection of aberrant speech using machine learning function descriptors and classifiers. Deep learning with feature descriptors and time frequency images is a better option. The majority of deep learning frameworks for speech-language pathology use a binary classification model. A network that can identify accurate medical conditions is required to construct a hardware system. It is essential todo a serious examination of time-frequency analysis using advanced deep learning algorithms. Current research is focused on creating a non-invasive, dependable, andcomputationally expensive architecture for detecting multiclass laryngeal lesions. In a realistic scenario application, compare the performance of a fully linked network versus a completely collapsed deep learning voice denoiser network. Three alternative time- frequency picture corpora are generated in the noise reduction training example.For applying in a realistic environment, the capability of a fully-connected network and afully convolutional deep-learning voice denoiser network is initially investigated. Denoised training samples are used to create three different time-frequency image corpuses. These multivariate image datasets are used to train three upgraded forms of the state-of-the-art convolution neuron network model using a 3D convolution kernel.

Keywords: Deep neural network , pathology classification , non invasive , random forest algorithm , speech language pathology , CNN algorithm , binary classification model voice dataset , pre processing , vocal issues , detection or identification.

I. INTRODUCTION

Different feature mining and machine learning methods were used to achieves a significant amount of effort has been done to distinguish a voice as natural or deviant. There are fewer literatures that have classified illnesses into subclasses. The current database, Voice ICar fEDerico II (VOICEDATASET), has a large number of vocal recordings from both healthy people and people with three different types of speech disorders. It is anticipated that a multi-class identification can be performed because no study has been completed on this dataset to our knowledge. Using a 3D convolution kernel, these multimodal picture datasets are utilisedto train three Improve three latest neural network models. Are using the "Group Decision Analogy" method to train and acquire the global maximum for current classification challenges. The idea of group selection analogies was inspired by well clustering and optimization algorithms algorithms. We get through the three stages to enhance the expected score. In order to recognise good hyperkinetic breathlessness, hypokinetic dyspnea, and reflux soft tissue infections in first layer, these extended deep learning networks in three datasets. Stages 2 and 3 of the prognosis are then given. Before using the Group Decision Analogy, a score of 80.59 percentage will be obtained, which will later be enhanced to97.7%. 98 percent of the time, hypokinetic sound and reflux laryngitis were diagnosed. Utilizing machine learning function descriptions and classifiers, several references supportthe non-invasive identification of abnormal speech. A preferable solution is deep learning with descriptors , time and frequency images.

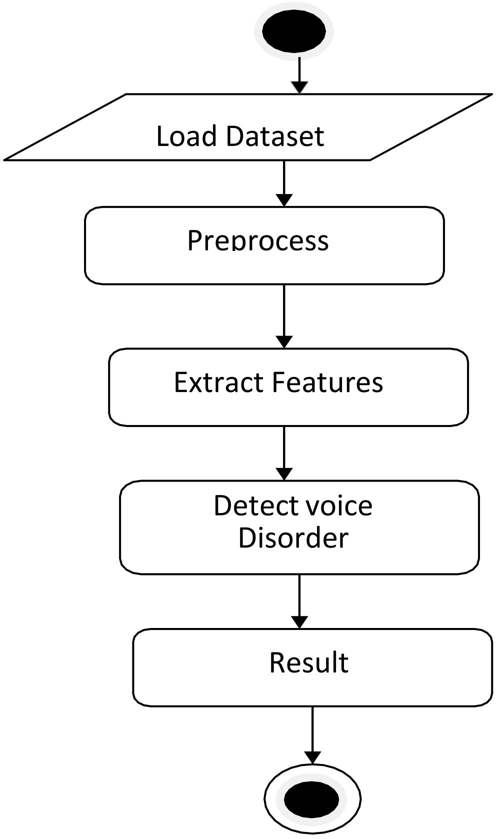
II. RELATED WORK

Through Feature engineering and machine learning approaches are used for a number of ways., a significant amount of effort has been done to distinguish a voice as natural or deviant. A category classification of diseases has been done in fewer literatures. The present dataset, Voice ICar fEDerico II (VOICEDATASET), has a large number of healthy vocal samples as well as three distinct voice collectives' disorders. Is anticipated which a multi class categorization possibly be performed, this dataset has not been subjected to any study toour knowledge. The current article presents a system architecture for pathological voice detection that offers excellent, per-class efficiency for three common voice diseases. A deep learning denoiser architecture is used as a pre-processing phase to delete non-speech artefacts. From a speech signals. To investigate their individual performance, a completely linked network and a fully connected network CN framework of deep learning are separately deployed as the main operator of denoiser. Second, it is well acknowledged that when compared to feature vectors, deep learning produces astonishingly good outcomes withcolour photos. With a

DOI: 10.17148/IJARCCE.2023.12224

deep learning network, researchers are investigating whether the time frequency visual patterns of voice data might impact accurate multiclass classification. The proposed system has the end to end framework with efficient hybrid learning design to classifying diseases in voice signals combining signals to enhance accuracy, and prediction-score filtering techniques. A description of data and its pre-processing is provided first, followed by a detailed stage process architecture methodology. The VOICED database is being utilised to show the proposed architecture's functionality.

III. PROPOSED METHODOLOGY



A. Short Time Fourier Transform (STFT)

It is a Wave equation transform that is used to calculate a signal's sinusoidal perceptible that changes over time. In practise, the STFT is calculated by dividing the lengthy signals into short segments of equal duration and computing the Fourier transform of every segment separately. This exposes each short segment's Fourier spectrum. Then, as a function of time, plot the shifting spectrum. A spectrogram, often known as a waterfall plot, is a type of spectrum display that is commonly used in SDR (Software Defined Radio) systems. A 2 24- point Fast Fourier Transform (FFT) on a computer is often used for fulband width displays that cover the complet spectrum.

B. Convolutional Neural Network (CNN)

Constraints a type of neural network that uses is a deep learning method whitch takes an inputpicture and assigns weights and biases to various regions / items in the picture.allowing them to be recognized. ConvNet requires considerably less pre-processing than other classificationalgorithms. ConvNet is smart enough to become familiar with these filters. / attributes VisualCortex, whereas the primitive technique requires well-trained and manually filters. Individual neurons can only respond to stimuli in a specific portion of the field of vision called a receptive field.

C. Data Preprocessing

We begin by looking at the columns that contain missing data. We see that, while many of these columns have missing values, the majority of these missing values correspond to mistakes linked with attribute values. We also see that two error properties - the positive and negative errors associated with Equilibrium Temperature - have missing values. As a

DOI: 10.17148/IJARCCE.2023.12224

result, we decide to remove them entirely because their values cannot be imputed. We see a severelyskewed distribution of a few error characteristics for the remaining error attributes. It would be unwise to replace these numbers with their average, therefore we opt to fill in the gaps with the median error value the VOICED record has been chosen. It was gathered in the medical room of the "High Performance Computing and Networking Institute of the Italian National Research Council (ICARCNR)" in partnership with the Naples University Hospital "Federico II" by Ugo Sesari and a team.

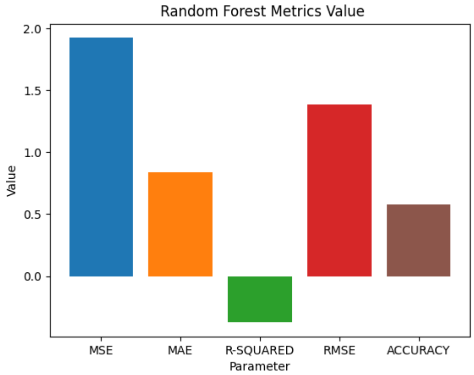
The vote was recorded in the Waveform Database (WFDB). All samples contain an "a" vowel signal that is 4.76 seconds long. This dataset contains 151 pathological and 57 healthy audio samples, with a total of 208 adult audio samples from the age group that contributed to 1870 for women (98 pathology, 37 health). 73 are male (52 pathological, 21 healthy). 151 morbid voices can be divided into three categories, 72 samples of hyperkinetic dysphonia, 41samples of hypokinetic dysphonia and 38 samples of reflux esophagitis.

D. Feature Extraction

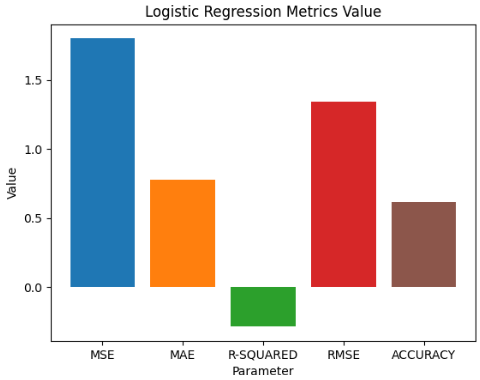
We are extracting the features from the voice dataset using MFCC. Windowing signals, applying DFTs, and obtaining size logs are all examples of MFCC feature extraction approaches. Then, on the Mel scale, distort the frequency, followed by DFT in reverse. MFCC is a cepstrum coefficient calculated using a warped frequency spectrum centred on human hearing. When determining the MFCC. The audio signal is first windowed, and thenit is separated into frames. The Mel Frequency Cepstrum Coefficient (MFCC) was first developed for detecting monosyllabic words in consecutively uttered phrases, but not for identifying the speaker.

IV. RESULTS AND DISCUSSION

In our project we are mainly consetrating on three different types of voice infections they are Hyperkinetic dysphonia Voice treatment significantly improves the voice of adults andchildren with hypermotor abnormal diseases with prenodular lesions and soft nodules, impacting a range of acoustic characteristics. Hypokinetic Hoarseness is a kind of dysphonia that happens at the glottic level. In the case of voice development problems, the entire organism and all of the systems involved in voice formation must be investigated. In this instance, the injured component of the vocalisation system may be identified and treated effectively disorder.



Random Forest Algorithm Results.



Logistic Regression Algorithm Results

DOI: 10.17148/IJARCCE.2023.12224

V. CONCLUSION

At last our project will compare the input data over the trained data and show the highly matched data as the results and tells whether the patient have any voice pathology or disorder.

This project can be used in remote villages where they don't have travel for high tech hospital for examine their disorder its cost effective and saves time.

REFERENCES

[1] J. R. Orozco-Arroyave et al., "Characterization methods for the detection of multiple voice disorders: Neurological, functional, and laryngeal diseases," IEEE J. Biomed. Health Informat., vol. 19, no. 6, pp. 1820–1828, Nov. 2015, doi: 10.1109/JBHI.2015.2467375.

[2] G. Gidaye, J. Nirmal, K. Ezzine, and M. Frikha, "Wavelet subband features for voice disorder detection and classification," Multimedia Tools Appl., vol. 79, nos. 39–40, pp. 28499–28523, Oct. 2020, doi: 10.1007/s11042-020-09424-1.

[3] I. Hammami, L. Salhi, and S. Labidi, "Voice pathologies classification and detection using EMD- DWT analysis based on higher order statistic features," IRBM, vol. 41, no. 3, pp. 161–171, Jun. 2020, doi: 10.1016/j.irbm.2019.11.004.

[4] A. Al-nasheri, G. Muhammad, M. Alsulaiman, and Z. Ali, "Investigation of voice pathology detection and classification on different frequency regions using correlation functions," J. Voice, vol. 31, no. 1, pp. 3–15, Jan. 2017, doi: 10.1016/j.jvoice.2016.01.014.

[5] N. Steffen, V. P. Vieira, R. K. Yazaki, and P. Pontes, "Modifications of vestibular fold shape fromrespiration to phonation in unilateral vocal fold paralysis," J. Voice, vol. 25, no. 1, pp. 111–113, Jan. 2011, doi: 10.1016/j.jvoice.2009.05.001.