**AUTOMATED DISEASE PREDICTION FROM VOICE MESSAGES USING MACHINE LEARNING TECHNIQUES**

Lalit Kumar Yadav

Department of Mathematics, Amrita Vishwa Vidyapeetham, Tamil Nadu, India

lalitkumar.ndit@gmail.com

**ARTICLE INFO**

**Keywords:**

MFCC

Spectrogram

Neural Network

Long Short Term Memory

Bi-LSTM

Random Forest Model

**ABSTRACT**

In the recent years there has been a rapid change in the way that technology used in betterment of human life. The field of Medicine has also progressed like using the Computer Vision, Image Processing and medical imaging technology and one such advancement is the detection of the diseases by Voice. Researchers have shown that assessment of voice disorder is important for tracking the clinical patient’s conditions and early detection of high-risk diseases.

In this paper I propose detection and prediction of diseases with the help of Machine Learning algorithms like LSTM, Bi-LSTM, and Random Forest. Experimental Results show that the proposed method has the ability to closely predict the result for the diseases.

1. **INTRODUCTION**

Most of the today’s population like to live in urban cities since the cities have better health infrastructure and other benefits, but there’s always a portion of people who live in Village or in Rural area, and the remote area people don’t generally have proper access to the medical facilities and are not very much aware of the various disorders of the voice and the hospital is an expensive place .So we are trying to develop a system which can give an idea of disease by just voice samples.

With the advancement of technology today we are able to at least get an idea of the disease we are suffering from. This is done by the Speech Recognition models/Systems (SRS) but these technologies ignore the situation a person is in by illness which causes slurred and incomprehensible speech. There are various diseases like cough [1], Parkinson [2], schizophrenia [3] etc. The Symptoms of schizophrenia can be assessed quantitatively through several scales such as the scale for Assessment of Thought, Language and Communication, Positive and Negative Symptom Assessment and so on. This assessment requires a lengthy conversation between the health professional and the patient and is time-consuming. The disease are a naturally occurring process so we have no control over it but we can take all necessary measures for it but first step of measure is the detection of it.

Speech Recognition is helpful in the medical sector because it can drastically reduce the time required to manually fill and hence enabling a smoother and longer patient-doctor time. While most Voice/Speech Recognition filters out the background noise there’s always a risk of background sound of instruments. This Speech Recognition or Voice Recognition is fastly being adopted in many sectors and according to a study around 70% [4] of medical sector are currently using it.

We need to understand the human speech production process to deal with disease prediction professionally. Phonetics is the study of the sounds produced by human speech. There are two types of phonetics – Acoustic and Articulatory. Acoustic phonetics is the subfield of linguistics which focuses on the acoustic and physical aspects of sound waves, while the articulatory phonetic is the method through which human bodies are employed to speak.

Since the speech of different people with different accent, style of speaking are different, detecting diseases is a challenge. A small part of speech signal of an ill person might sound normal to normal speech of another person with normal accent. Representation of speech for processing is not yet fully matured to easily allow such differentiation particular when using it to train deep learning or machine learning models.

This Paper focuses on comparing various models for predicting and detecting disease from the voice of the patient/person. The main purpose is to check if Machine Learning Models can be used to validate the dataset to get medical findings or issues. We are applying ANN (Artificial Neural Network), LSTM (Long Short Term Memory), BI-LSTM (Bi-Directional Long Short Term Memory), Random Forest and Gradient Boosting Methods on the extracted features from the dataset. For this paper I am using dataset consisting of 6661 wav file and one csv file. The Basic Pre-processing is done on the dataset and features are extracted from the data. The acquired feature is displayed in terms of number of words in each category, the length of phrases and the number of words in each medical speech phrase.

The rest of the paper is structured as follows: Section 2 describes the methods used for this study. Section 3 summarizes the available dataset for prediction purpose and the platform and libraries, dataset collection, pre-processing, feature extraction and the evaluation parameters. Section 4 includes the proposed work’s experimental findings and analyses. Section 5 contains the study’s discussion and closes the paper with an explanation of the acquired results and their comparison to baseline techniques with room for improvement.

Train the Model

Pre-processing

Result

Applying ML Algorithms

Testing the model

Fig 1: Schematic Diagram For Disease Prediction

As for the research for this work we have gone through various articles and papers.[4] Used simple linear classifiers trained on the TORGO database to detect Dysarthria. They converted the input signal to time domain filter banks and passed through LSTM and fully connected layers in the model and the hyper parameters tuned to get success rate of 72% on TORGO. [5] Used the Convolutional Neural Network for training the model through a softmax classifier with a success rate of 68%. [6] Used a unidirectional recurrent neural network (RNN) of 1D convolutional layer of window length of 15, stride length of 4 and 196 filters on OpenSLR LibriSpeech SLR 12 corpus.

Kristiawan[7] proposed a data augmentation strategy using Adding white noise, pitch shifting, time stretching which is processed using DA-DNN7L.The data Augmentation approach is used to increase the limited data quantity of Indonesian ethnic speakers[8],while the seven layer DNN gives the best accuracy of 99%,loss of 0.05 and addition of 400 data. Niousha [9] were trying to detect Parkinson disease in patients. For this they have used several fuzzy classifiers like Inductive Fuzzy Classifier, Fuzzy Rough classifier on the “Parkinson Speech Dataset with multiple types of sound recordings” [10]. It concluded that the FURIA, MLP-Bagging-SGD performed the best.

David, Kiss, Tulics [11] used a LSTM auto encoder hybrid with multi-task learning as input feature. They applied different speech databases

Problem Statement:

Many of the today’s machines learning models in the health care analysis are concentrating on one disease per analysis. For example one is for diabetes, other is for heart analysis, lung analysis etc. If a user wants to predict more than one disease, the person has to go through different sites. There is no common system where one analysis can perform more than one disease prediction for the people. Hence we are trying to get a better accuracy to predict the disease.

**Section 2**

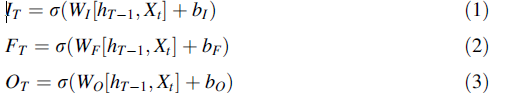
For this work, we have decided to take up Deep and machine learning models like NN (Neural Network), Random Forest, SVM (Support Vector Machine),LSTM(Long Short Term Memory).

A Neural Network is network of biological neurons (human brain like) or a network of nodes. In other words it’s a series of algorithm’s that endeavours to recognise underlying relationship in a set of data through a process that mimics the human brain. The Neural network has connection to the Artificial Neural Network. Neural networks with several process layers are known as deep networks and are used for deep learning algorithms. A neural network has Input layer, hidden layer and output layer. A neuron in neural network is a mathematical function that collects and classifies information according to specific architecture. There are various types of NN like Feedforward neural network, recurrent neural network, and convolutional neural network.

Support vector machine (SVM) is one of the algorithms in supervised machine learning. It’s used for both classification and regression problems. The main aim of the SVM algorithm is to create a best fit decision boundary that can separate n-dimensional space into classes so that we can easily put the new data point in the correct category in future. There are linear and non-linear SVM.SVM algorithm can be used in face detection, image classification, text categorization etc. The working of SVM is such that we have 2 tags for a dataset we try to make the support vectors properly classify the tags and to maximize the margin.

The LSTM(Long-short term Memory)architecture is a type of RNN(Recurrent Neural Network) used in deep learning. LSTMs have feedback connections, unlike traditional feed forward neural networks. It can process not only single data points (such as photos),but also entire data streams(such as speech and video).

A typical LSTM consists of four components: a cell, an input gate, and a forget gate. The equations are used to evaluate all of them Eqs.(1,2,3) The three gates control the entry and outflow of information, and the cell holds data across arbitrary time intervals. LSTM networks are ideally suited for categorizing, analysing, forecasting time series data as there may be unpredictable delays between significant occurrences in a time series.LSTMs were created to address the vanishing gradient problem that can occur when standard RNNs are trained.

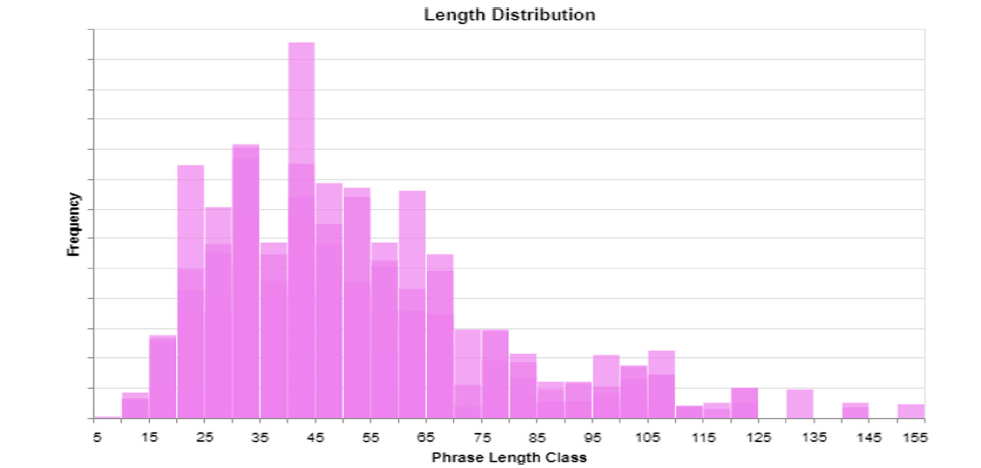


Here It input gate, Ft forget gate, Ot output gate,Wx is weight of all gates,X neurons, ht-1 output of previous LSTM block at timestamp(T-1),Xt input at current timestamp and bx biases for respective gates.

**Section 3**

For this project, we decided to take up the dataset of Medical Speech Utterance, Transcription and intent dataset [8]. The dataset consist of 8.5 hours of audio utterances paired with text for common medical symptoms. This dataset contains both audio and corresponding transcriptions. It contains medical symptoms like knee pain or headache .Each utterance was created by individual human contributors based on given symptom. It has 6661 wav files and 1 csv files which give an over-view of the recordings. For the train and test the data was split with a ratio of 70:30, 80:20. For this project the main libraries which were used was the librosa, vaex, pandas and FastAI and pre-trained models.

For the pre-processing step we did the basic pre-processing i.e. removing Nan values and removing duplicates if present any, tokenization, stemming, lemmatization etc. for the csv file and noise reduction was also applied on the voice samples using noise reduction

Forward backward filter was used to eliminate noise from wave file. Trimming was also done on the voice samples.

As we know the most important step in the data analysis is the selection of features from the data. The most critical component of feature extraction is to identifying the aspects that are most relevant to the issue statements. In this work, MFCC (Mel frequency cepstral coefficient) features were extracted. These features are highly related to human being experiencing pain, and were chosen based on experiments conducted on speech.

The feature selection method used in this study was to filter feature groups before giving them to neural networks (NN).

For the evaluation we have taken up accuracy and loss as the parameter. Accuracy is a metric that’s used to evaluate which model is highly effective in identifying connections and patterns among variables in a dataset using the input or training data. It’s straightforward to compute by dividing correct guesses by the total number of forecasts.

Accuracy=(TP+TN)/(TP+TN+FP+FN)

Where TP,TN,FP,FN are True Positive, True Negative, False Positive, False Negative respectively.

Loss is other parameter used for the evaluation. Loss is the number which tells how wrong the model predicts. It’s a kind of penalty for wrong prediction.