ISSN No:-2456-2165

Malayalam Speech Recognition

Abhirami K
Dept. of Computer Science and Engineering
Sahrdaya College of Engineering and Technology
Kodakara, India

Aiswarya P S
Dept. of Computer Science and Engineering
Sahrdaya College of Engineering and Technology
Kodakara, India

Aishwariya D P
Dept. of Computer Science and Engineering
Sahrdaya College of Engineering and Technology
Kodakara,India

Abstract:- The project is based on the development of state-of-the-art large vocabulary continuous speech recognition (LVCSR) system for the Malayalam language. Problems of existing speech recognition are lack of accuracy and misinterpretation, time cost and productivity, accents and speech recognition, background noise interference. The simulation of human intelligence in computers refers to artificial intelligence (AI) which includes Machine Learning, Natural Language Processing, Computer Vision and Robotics. In audio files or video files that are large and have minutes in length, many files have a variety of audio and audio files. In this project, transfer flow technique is used.So the aim of the proposed system of speech recognition is to collect thousands of datasets of each category irrespective of their gender and also they can be of any age group and train them according to their native sequence so as to increase the accuracy level.

Keywords:- *Artificial Intelligence; Machine Learning; Large Vocabulary Continuous Speech Recognition; Support Vector Machine, Tensor Flow.*

I. INTRODUCTION

Speech is a simple and usable technique of communication between humans, but nowadays humans aren't limited to connecting but even to the different machines in our lives. The most important is the computer. So, this communication technique is often used between computers and humans. This interaction is completed through interfaces, this area is called Human-Computer Interaction (HCI).Presently, computers have already replaced a tremendous number of humans in many creative professions. Speech recognition can be predicted using a computer. Our project focuses on the development of state-of-the-art large vocabulary continuous speech recognition (LVCSR) systems for the Malayalam language. We choose to listen to the desired sound from a large file. Here machine learning is used to classify the speech. Machine learning is to make computers able to learn problems and solve problems on their own. In our project the model is created using tensor flow. Tensor flow is one machine learning approach. It is open source. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that let researchers push the state-of-the-art in Machine Learning and developers can easily build and deploy Machine Learning powered applications. Here we have also used a transfer Learning approach. A transfer learning method is used to transfer two types of knowledge to different datasets.

A. Motivation

The state of Kerala has 14 revenue districts. Each district has its own way of speaking Malayalam which is the mother tongue of the state. The state is also popularly called as land of backwaters which makes it a spot for tourists. Even though Malayalam is spoken all over the state, it's occasionally hard for people to recognise the language. Hence with our software one could recognise the slang being used.

B. Proposed system

The Malayalam language is a language that people use in different slangs. People from each part of Kerala use a different slang. For each common word people from each region use different slang to pronounce it or say something completely different. The Google assistant which we could revoke by saying OK Google is used to translate words and sentences of different languages to the language we request. It can't understand or differentiate different slangs or the words that are used by the people throughout these regions. It's often misunderstood by Google Assistant on the words used by the people. The accuracy hence is low. The problem of existing speech recognition is a lack of accuracy and misinterpretation, time cost and productivity, accents and speech recognition, background noise interference. Here the transfer learning approach is used.

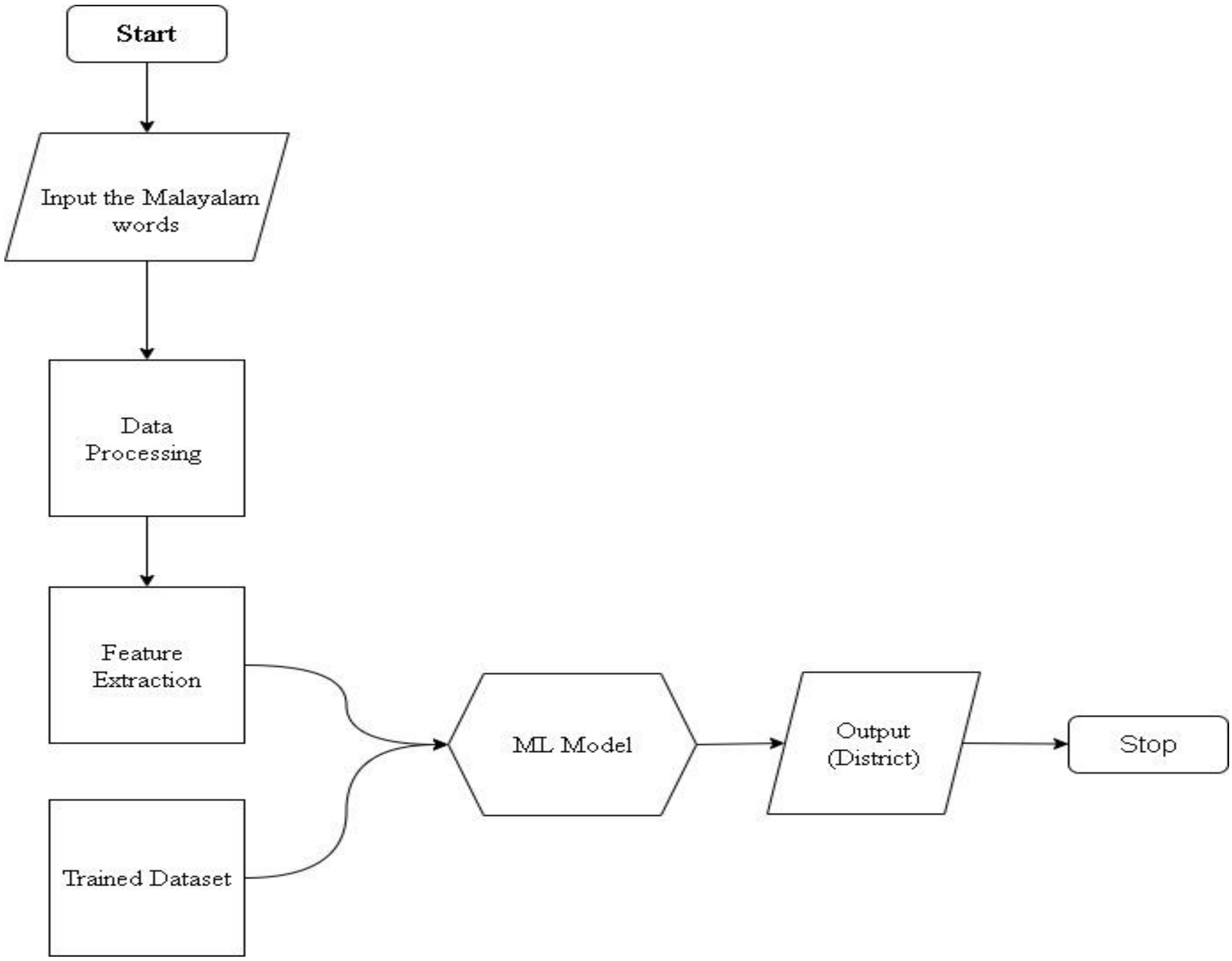

FIG 1

II. METHODOLOGY

The project was implemented using google cloud platform. The Google cloud platform is a suite for cloud computing services provided by Google. It runs on the same infrastructure that Google uses internally for its end user's products such as Google search, File Manager, Gmail, Classroom et cetera. It provides modular cloud services including computing, data storage, analytics and machine learning.It provides infrastructure as a service ,serverless computing environment, and platform as service. It is part of google cloud service.

The whole project was also done on TeachableMachine as a trial for checking the credibility of the project and for also getting a more wide picture of the project's working.It is also a product of Google which actually helps anyone who is interested in learning and conducting experiments or doing any projects related to AI. It helps one train their computer to recognise images, audios and also helps in creating snake games.

The audio classification problem is used here to classify the slangs. It is also called acoustic event detection. It is the process of listening and analysing the audio recordings. For classification, audio recordings are converted into spectrograms, then we input them to a CNN plus linear classifier model which produces predictions of the class in which they belong. The simplest and easiest way to explain the process of audio classification is that: firstly the audio is collected; it is then converted into the form of a spectrogram. Later it is then inputted to CNN architecture.After further procedures it is then inputted to linear classifier which helps in classifying those audio into its respective labels or classes. It requires understanding of the underlying frequency structure of acoustic signals. Here we need to build a model which has knowledge of features for each audio class so that during the evaluation, it understands and classifies a given audio segment into corresponding class.

BERT is a natural language processing model proposed by researchers at Google Research 2018. The main reason for the good performance of the BERT model is the use of semi-supervised learning. The model is trained for a specific task enabling it to understand the pattern of the language.It has language processing capabilities that can be used to empower other models. It is basically an encoder stack of transformer architecture. An encoder-decoder network which uses self attention on the deoder side. Long-short term memory is an artificial recurrent neural network architecture used in deep learning.

III. RESULT

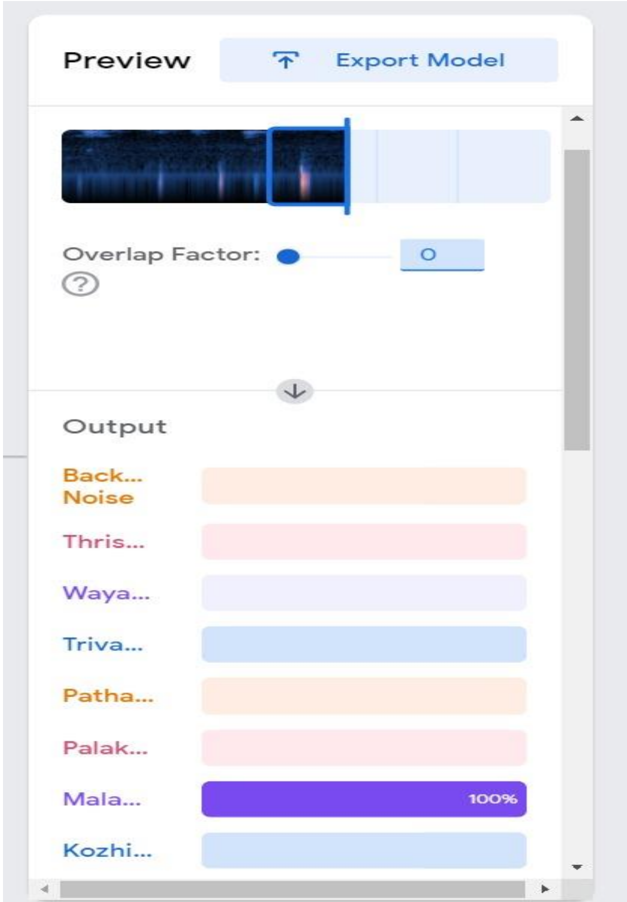

FIG 2: OUTPUT AFTER TRAINING THE MODEL

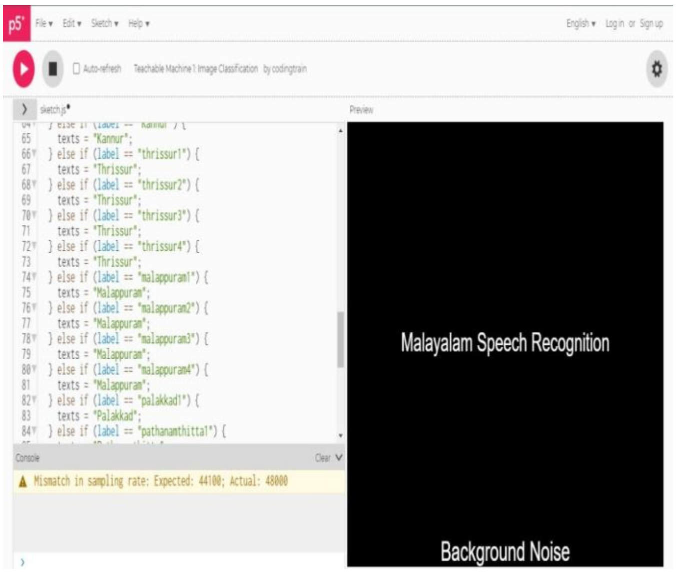

FIG 3: AFTER DEPLOYING THE MODEL

IV. CONCLUSION

In this work, we proposed a new method for malayalam speech recognition. Our system automatically detects the district. Our system uses the bert model. Model is trained in the gcp. The application is more important to the people who can't understand the slang. After creating the model we achieved an accuracy of 70%.

FUTURE WORK

Now we have created our model using the bert model. In the future we will create our model using the mum model. Also the dataset was less. We will collect more voices, so that we can increase the accuracy.

REFERENCES

[1]. Yuki Takashima, Ryoichi Takashima, Tetsuya Takiguch and Yasuo uriki, ,"Knowledge transferability between the data of the persons with Dysarthria speaking different languages of dysarthric speech recognition", IEEE Access, vol. 7, pp. 164320 - 164326, April 2020.

[2]. G. E. Dahl, D. Yu, L. Deng and A. Acero, "Large vocabulary continuous speech recognition with context-dependent DBN-HMMs", Proc. IEEE Int. Conf. Acoust. Speech Signal Process., pp. 4688-4691, May 2011.

[3]. S. Chandrakala and N. Rajeswari, "Representation learning based speech assistive system for persons with dysarthria",IEEE Trans. Neural Syst Rehabil Eng, Vol. 25, pp. 1510-1517, Sep. 2017.

[4]. S. J. Pan and Q. Yang,, "A survey on transfer learning", IEEE Trans.Knowl.Data Eng, vol. 22, pp. 1345-1359, Oct. 2010.

[5]. J. Duffy, Motor Speech Disorders: Substrates Differential Diagnosis and Management, New York, NY, USA:Elsevier, 2013.