iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VIII **Month of publication:** August 2021

DOI: https://doi.org/10.22214/ijraset.2021.37296

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

Keyword Recognition Device Cloud Based

Apurv Singh Yadav[1], Abhi Yadav[2], Bhavya Malhotra[3], Aman Shah[4]
[1, 2, 3, 4]*Electronics and Communication Engineering, ABES Engineering College Ghaziabad, UttarPradesh, India*

Abstract: Over the past few decades speech recognition has been researched and developed tremendously. However in the past few years use of the Internet of things has been significantly increased and with it the essence of efficient speech recognition is beneficial more than ever. With the significant improvement in Machine Learning and Deep learning, speech recognition has become more efficient and applicable. This paper focuses on developing an efficient Speech recognition system using Deep Learning.

Keywords: speech recognition, deep learning, internet of things

I. INTRODUCTION

Speech is the most important technique of communication among humans. The ability to correctly recognize speech by a machine is known as automatic speech recognition. The main objective of this paper is to develop a highly efficient model for speech recognition using deep learning considering its further use in personal devices and applications.

Speech recognition is a fundamental part of speech processing and therefore in today's world finds whole new importance in everyday tasks. For Example saying "Ok Google" or "Alexa" before searching anything over the web on respective platforms. By this model we aim to achieve maximum efficiency so as to be used in applications with least errors.

The research on speech recognition can be dated back to 1964 and has been fundamentally developed. In our model we have used MFCC(mel-frequency cepstral coefficients) with feature extraction using Deep learning methods mainly CNN(Convolutional Neural Networks). In our System we try to achieve the results for the following steps with maximum accuracy and efficiency so as to build a whole efficient system :

1) Collecting Data from clients or users with least noise in it.
2) Training Model with existing dataset.
3) Testing new data from clients to get results as well as continue training the model.
4) Working on processed data and its further deployment back to the user.

This paper is not concerned with the research on automatic speech recognition rather on developing the most efficient system with the current technology used and its accessibility in IOT(internet of things).

II. NEED FOR DEVELOPMENT

During the past decade, there has been tremendous developments in technology and its use in everyday life. The Internet of things is one the things that plays a huge role in this development and it will continue to do so.

The main aspect of Speech recognition is to make machines Voice Powered or have nearly whole voice enabled interaction. As the Internet of Things (IOT) is developing everyday significantly, speech recognition will be a very important feature and development to the applications and the industry. The application of speech recognition would make tasks significantly easier and viable to people with disabilities also, therefore there is a need for development in speech recognition and analysis to achieve the desired results quickly and efficiently.

III. LITERATURE REVIEW

A. Overview and Discussion

Aim of this project is develop an efficient Speech Recognition system that can initially identify keywords and also can be expanded to more practical applications in future. Speech recognition is a part of Machine Learning which involves Deep Learning to train our proposed model on real-time data and come up with a system that can accurately predict the keyword from the user.

Deep Learning is a machine learning technique that constructs artificial neural networks to mimic the structure and function of the human brain. In practice, deep learning, also known as deep structured learning or hierarchical learning, uses a large number of hidden layers -typically more than 6 but often much higher - of nonlinear processing to extract features from data and transform the data into different levels of abstraction (representations).To achieve this complicated task we has used Python as our scripting Language and Have Imported :

©IJRASET: All Rights are Reserved

Librosa which is a python package for music and audio analysis. It provides the building blocks necessary to create music information retrieval systems and implements MFCC function to extract features from our audio data
OS provides a portable way of using operating system dependent functionality. We can build an interactive system using it.
TensorFlow is an open source library for numerical computation and large-scale machine learning. It bundles together a slew of machine learning and deep learning models and algorithms and makes them useful.
Keras which is a TensorFlow API that has implemented functions of Deep Learning algorithms

B. Preparation of dataset

A data set is needed to train and check the efficiency of our model. We are using "Speech Commands Dataset" , a free data service provided by Google. A speech signal needs to be processed into some digital form so that a computer can understand it . For that we are extracting MFFCs from input audio.
MFCC: MFCCs are the Mel Frequency Cepstral . MFCC takes into account human perception for sensitivity at appropriate frequencies by converting the conventional frequency to Mel Scale, and are thus suitable for speech recognition tasks quite well (as they are suitable for understanding humans and the frequency at which humans speak/utter).We have used 13 Mel Frequency coefficients into consideration as features when training models.

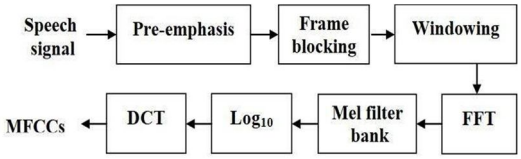


Fig. MFCC WORKING

We are using Librosa which is a python package that offers a function MFCC() with which we can extract Mel Frequency Cepstral Coefficients and process that data as per our requirement.

C. Load Train/Validation/Test Data Splits

Processed data now is divided into three parts that is Train, Validation and test for training of our proposed model.
Training is data we feed and expect its identification. Validation dataset is a sample of data held back from training our model that is used to give an estimate of model skill while tuning model's hyperparameters.
Test is the expected data that is compared with our processed training data and we analyse our model based on that

D. Building an Efficient Model

Since currently our project is limited to keyword detection we need a coherent Deep learning algorithm which can be precise and efficient. CNN (Convolutional Neural Network) has the property of backpropagation and it's output data can also be changed to be one-dimensional, allowing it to develop an internal representation of a one-dimensional sequence. For implementing our CNN model we have used Tensorflow as it provides a deep learning library and Keras is its API with which we can build various complex models for future expansion. Our CNN model will be using 3 layers to detect the input keyword. First layer of CNN we have a regularization technique which makes slight modifications to the learning algorithm such that the model generalizes better.
Second and Third layer does batch normalization and downsamples the output. To further improve our model we have used 2 dense layers as a densely connected layer provides learning features from all the combinations of the features of the previous layer and adds a lot more stability to our system.

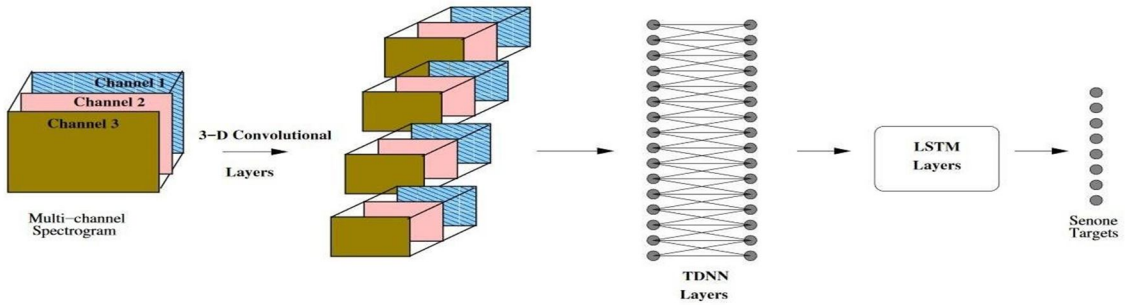


Fig. Blocked Schematic of 3 layer CNN Architecture used for Speech Recognition

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

E. Training the Model

For training the model Keras provides a fit() function that will process the data based on our provided parameters. We have set Epoch to 40 that is number of times our model will process a set of data repeatedly

F. Evaluating the Model

For evaluation we have defined Test Error and Test Accuracy. Test Error evaluates training and test data set difference which will give Test Accuracy that shows how accurate our model is

G. Saving the Model

Keras has save() function which will automatically save our trained model to h5 file and now we can use it for prediction

IV. CONCLUSIONS AND FUTURE SCOPE

A prediction data set was tested and our model is working at efficiency of approximately 95%.

```
3498/3498 [==============================] - 1s 237us/sample - loss: 0.2278 - accuracy: 0.9514
Test error: 0.22780332202704176, test accuracy: 0.9514008164405823
```

Fig. Result Observations

For the future if a more complex network is fed then it can detect keywords and sentences that are not fed into the model during training.

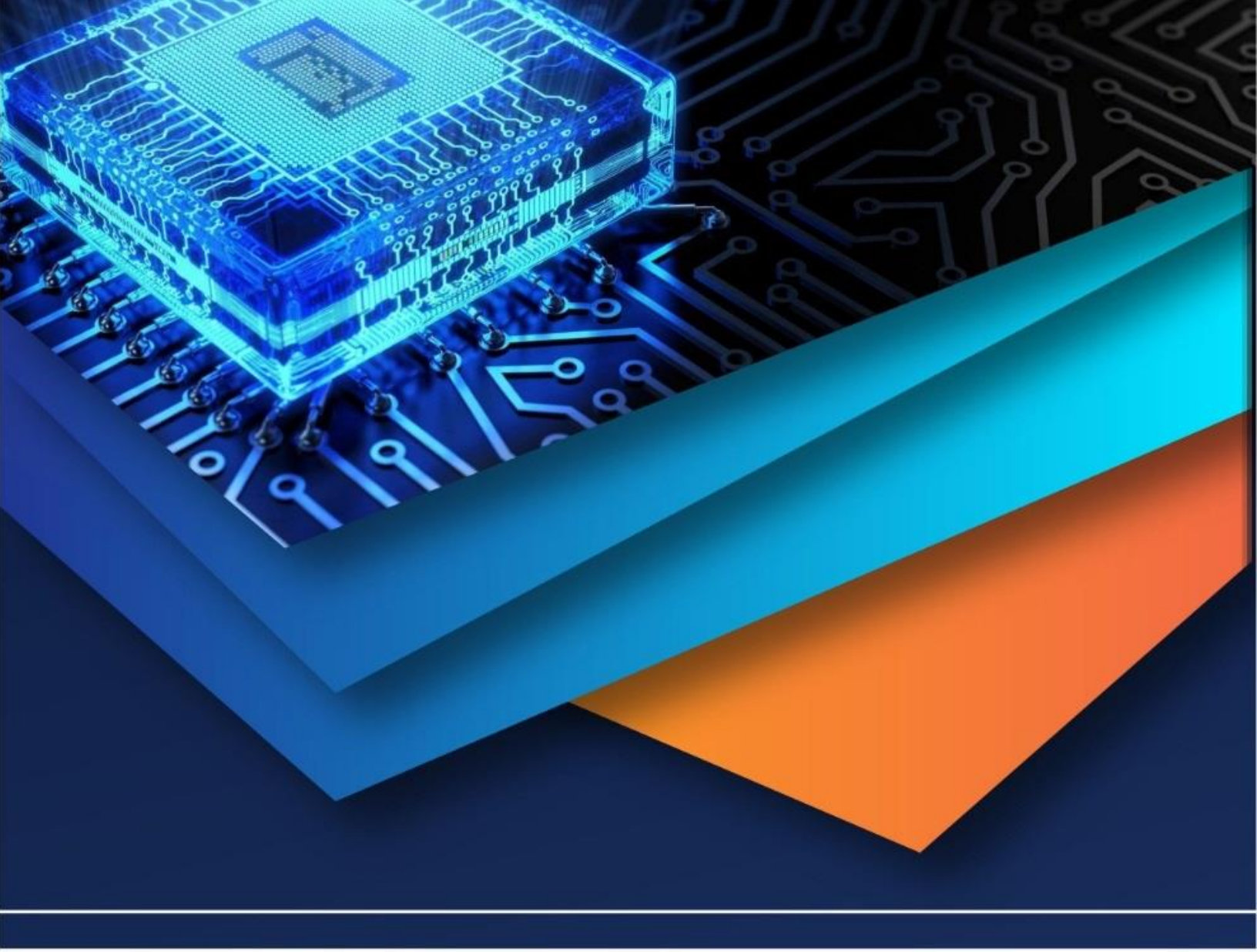
V. ACKNOWLEDGEMENTS

We would like to thank our guide and professors of the Electronics and Communication Engineering Department, ABES Engineering College, Ghaziabad For their consistent guidance, support and facilities extended to us.

REFERENCES

[1] https://ai.googleblog.com/2017/08/launching-speec h-commands-dataset.html
[2] Povey, Daniel & Ghoshal, Arnab & Boulianne, Gilles & Burget, Lukáš & Glembek, Ondrej & Goel, Nagendra & Hannemann, Mirko & Motlíček, Petr & Qian, Yanmin & Schwarz, Petr & Silovský, Jan & Stemmer, Georg & Vesel, Karel. (2011). The Kaldi speech recognition toolkit. IEEE 2011 Workshop on Automatic Speech Recognition and Understanding.
[3] Kamath, Uday & Liu, John & Whitaker, James. (2019). Deep Learning for NLP and Speech Recognition. 10.1007/978-3-030-14596-5.
[4] Saleh, Suaad & Ben Dalla, Llahm Omar. (2020). Deep Learning for Speech Recognition. 10.13140/RG.2.2.15714.15048.
[5] M. Mehrabani, S. Bangalore and B. Stern, "Personalized speech recognition for Internet of Things," 2015 IEEE 2nd World Forum on Internet of Things (WF-IoT), Milan, 2015, pp. 369-374, doi: 10.1109/WF-IoT.2015.7389082.
[6] Bai, Zhongxin & Zhang, Xiao-Lei & Chen, Jingdong. (2020). Speaker Recognition Based on Deep Learning: An Overview.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)