iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** V **Month of publication:** May 2020

DOI: http://doi.org/10.22214/ijraset.2020.5359

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 8 Issue V May 2020- Available at www.ijraset.com

Speech Recognition using Deep Neural Network Neural (DNN) and Deep Belief Network (DBN)

Prof. Kirti Rajadnya [1], Purva Ingle [2], Vinit Tawsalkar[3], Shivani Teli[4]
[1]*Prof.Kirti Rajadnya, Dept. of Information Technology Engineering, SSJCOE, Maharashtra, India*
[2] *Ingle Purva, Dept. of Information Technology Engineering, SSJCOE, Maharashtra, India*
[3]*Tawsalkar Vinit, Dept. of Information Technology Engineering, SSJCOE, Maharashtra, India*
[4]*Teli Shivani, Dept. of Information Technology Engineering, SSJCOE, Maharashtra, India*

Abstract: ***It is a difficult task of continuous automatic speech recognition, translating of spoken words into text due to the excessive viability in speech signals. In recent years speech recognition has been accomplishing pinnacle of success however it still has few limitations to overcome. Deep learning also known as representation learning or sometimes referred as unsupervised feature learning, is a subset of machine learning. Deep learning is becoming a conventional technology for speech recognition and has efficiently replaced Gaussian mixtures for speech recognition on a global scale. The predominant goal of this undertaking is to apply deep learning algorithms, together with Deep Neural Networks (DNN) and Deep Belief Networks (DBN), for automatic non-stop speech recognition.***

Keywords: Gaussian Mixture Model (GMM), Hidden Markov Models (HMMs), Deep Neural Networks (DNN), Deep Belief Networks (DBN).

I. INTRODUCTION

The task of continuous automatic speech recognition, translation of spoken words into text, still is a big challenge due to the high variability in speech signals. Such as, speakers may have different accents, pronunciations, or dialects, and speak in several different styles, at different rates, and in numerous emotional states. The presence of many environmental noise, resonance, different types of microphones and recording devices adds to the variability. Traditional speech recognition systems make use of Gaussian mixture model (GMM) based Hidden Markov Models (HMMs) to represent the sequential structure of speech signals. The reason to use HMMs in speech recognition is because a speech signal is usually observed as a piecewise stationary signal or a short-time stationary signal. The HMMs-based speech recognition systems are usually trained automatically, simple to use and computationally viable option. However, one of the major drawbacks of Gaussian mixture models is that they may be statistically inefficient for modeling records that lie on or close to a non-linear manifold in the data space. Deep Neural Network - Hidden Markov Model (DNN-HMM), has been proposed and broadly utilized in speech recognition. A Deep Neural Network (DNN), which is able to capture the underlying nonlinear relationship amongst data, is the conventional multi-layer perceptron with many layers, where training is usually initialized by using a pre-training algorithm. DNN-HMM alongside an unsupervised pre-training method to train a Deep Belief Network, which has the strong ability of feature learning provides a better recognition result. Due to this DNN-HMM with DBN the proposed model can achieve automatic continuous speech recognition which is more accurate and faster than traditional GMM-HMM system of speech recognition.

II. PROBLEM DEFINITION

To develop an efficient system which can automatically recognize speech, translate spoken words into text, using dataset by using Deep Neural Networks (DNN) based hidden Markov models (HMMs) and DBN to represent the sequential structure of speech signals, which is user friendly, fast and easy to use.

III. LITERATURE SURVEY

A. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [1] By W. Xiong, L. Wu, F. Alleva, J. Droppo , X. Huang, A. Stolcke. (Microsoft AI and Research.)

This paper represents the latest version of Microsoft's conversational speech recognition system for the Switchboard and CallHome domains. The system adds a CNN-BLSTM acoustic model to the set of model architectures combined previously, and includes character-based and dialog session aware LSTM language models in rescoring.

©IJRASET: All Rights are Reserved

B. A Multi-Task Learning Framework for Overcoming the Catastrophic Forgetting in Automatic Speech Recognition By - Jiabin Xue, Jiqing Han, Tieran Zheng, Xiang Gao, Jiaxing Guo. (Submitted on 17 Apr 2019).

In this paper, they plan to solve the CF problem with the lifelong learning and propose a completely unique multi-task learning (MTL) training framework for ASR. It considers reserving original knowledge and learning new knowledge as two independent tasks, respectively.

C. International Journal of Knowledge-based and Intelligent Engineering Systems, (21 March 2018) By Haridas, Arul Valiyavalappil, Marimuthu, Ramalatha, Sivakumar, Vaazi GAngadharan.

This paper states, a survey of speech recognition strategies suitable for human identification is discussed in this study. The main motivation of this survey is to explore the prevailing speech recognition strategies in order that the researchers can include all the required metrics in their works during this domain and the limitations in the existing ones can be overcome.

D. Speech Emotion Recognition Using Deep Neural Network Considering Verbal and Nonverbal Speech Sounds By Kun-Yi Huang; Chung-Hsien Wu; Qian-Bei Hong; Ming-Hsiang Su; Yi-Hsuan Chen.

This paper represents, Speech emotion recognition importance in conversation. It proposes a model that uses A Prosodic Phrase (PPh) auto-tagger was further employed to extract the verbal/nonverbal segments convolutional neural networks (CNNs). The proposed method achieved a detection accuracy of 52.00% outperforming the traditional methods.

E. Deep factorization for speech signal (2014) By Lantian Li, Dong Wang, Yixiang Chen, Ying Shi, Zhiyuan Tang, Thomas Fang Zheng.

This paper proposes a method in which the most significant factors are inferred firstly, and other less significant factors are inferred subsequently on the condition of the factors that have already been inferred. The limitation in Automatic Speech Recognition is the signals involve rich information, including linguistic content, speaker trait, emotion, channel and background noise, etc.

F. Leveraging automatic speech recognition in cochlear implants for improved speech intelligibility under reverberation (2018) By Shabnam Ghaffarzadegan, John H.L. Hansen Center for Robust Speech Systems (CRSS).

In this paper advantages of advances of DS in speech understanding for CI users in the presence of noise and/or reverberation in the forms of modified speech coding strategies or front-end signal enhancement are stated.

G. Listen, Attend and Spell (2015) By William Chan, Navdeep Jaitly, Quoc V. Le, Oriol Vinyals (Google Brain).

This paper tells about Listen, Attend and Spell (LAS), a neural network that learns to transcribe speech utterances to characters. The first component, the listener, is a pyramidal acoustic RNN encoder that transforms the input sequence into a high-level feature representation. The second component, the speller, is an RNN decoder that attends to the high-level features and spells out the transcript one character at a time.

IV. DESIGN AND IMPLEMENTATION

A. Hardware Requirements

1) Laptop



Figure 1: Laptop

The Laptop used here is with at least 8GB of ram and 1TB of HDD with Windows 10. It is mainly used to run the Deep Neural Networks (DNN). It is also training the Deep Belief Networks (DBN). All the complex data analysis and data load is handled using the python terminal and python packages. It captures the audio signals using a headset with microphone and performs the metrics.

©IJRASET: All Rights are Reserved

2) Headset with Microphone



Figure 2: Headset with Microphone

A headset with microphone is used because it converts sound waves into audio signal which is electrical. Due to this the speech recognition software gets audio input from the user.

B. Implementation

The flowchart is shown as below:

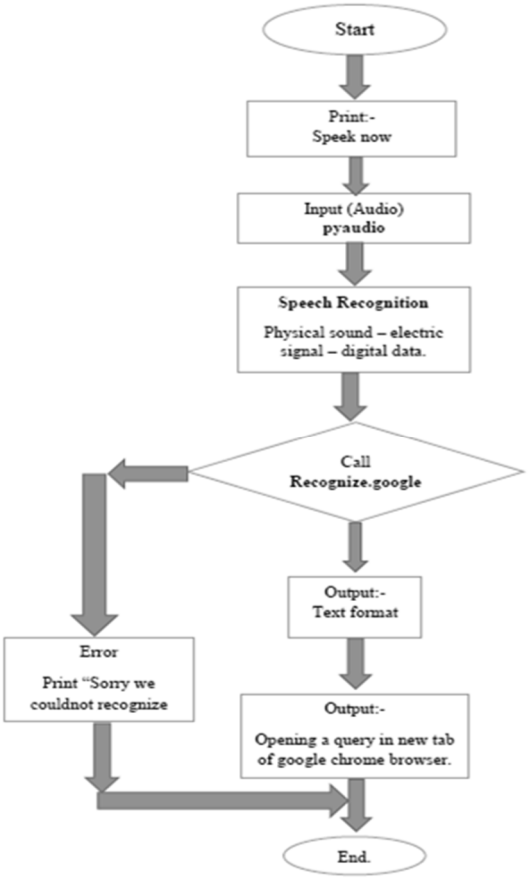


Figure 3: Flowchart of Speech Recognition System

1) Open a python-based GUI application.
2) Click on a button showed on application to take an audio input.
3) With a microphone source it will accept the audio.
4) Using recognize. Google, speech recognition package will convert audio input into text format.
5) If a recognizer is able to recognize the audio input then it will print as "Sorry we could not recognize your voice."
6) The recognized speech will be used in DNN-HMM and training DBN.

©IJRASET: All Rights are Reserved

The system architecture for Speech Recognition System using DNN-HMM and DBN is shown below:

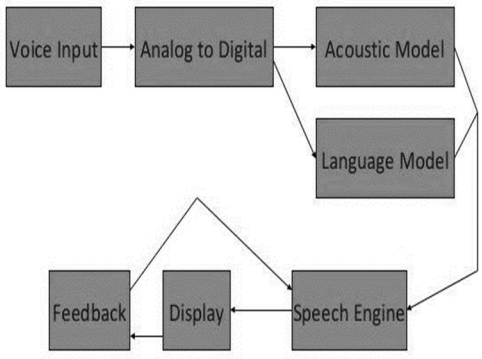


Figure 4: System Architecture of Speech Recognition System using DNN-HMM and DBN

C. Methodology

1) *Hidden Markov Model:* Modern general-purpose speech recognition systems are supported Hidden Markov Models. These are statistical models that output a sequence of symbols or quantities. HMMs are utilized in speech recognition because a speech signal is often viewed as a piecewise stationary signal or a short-time stationary signal. Speech are often thought of as a Markov model for several stochastic purposes.
2) *Neural Network:* Neural networks emerged as a beautiful acoustic modeling approach in ASR within the late 1980s. Since then, neural networks are utilized in many aspects of speech recognition like phoneme classification, isolated word recognition, audiovisual speech recognition, audiovisual speaker recognition and speaker adaptation.
3) *Deep Learning:* Deep Neural Networks and Denoising Autoencoders are also under investigation. A deep feedforward neural network (DNN) is a man-made neural network with multiple hidden layers of units between the input and output layers. Similar to shallow neural networks, DNNs can model complex non-linear relationships. DNN architectures generate compositional models, where extra layers enable composition of features from lower layers, giving an enormous learning capacity and thus the potential of modeling complex patterns of speech data.

V. RESULT

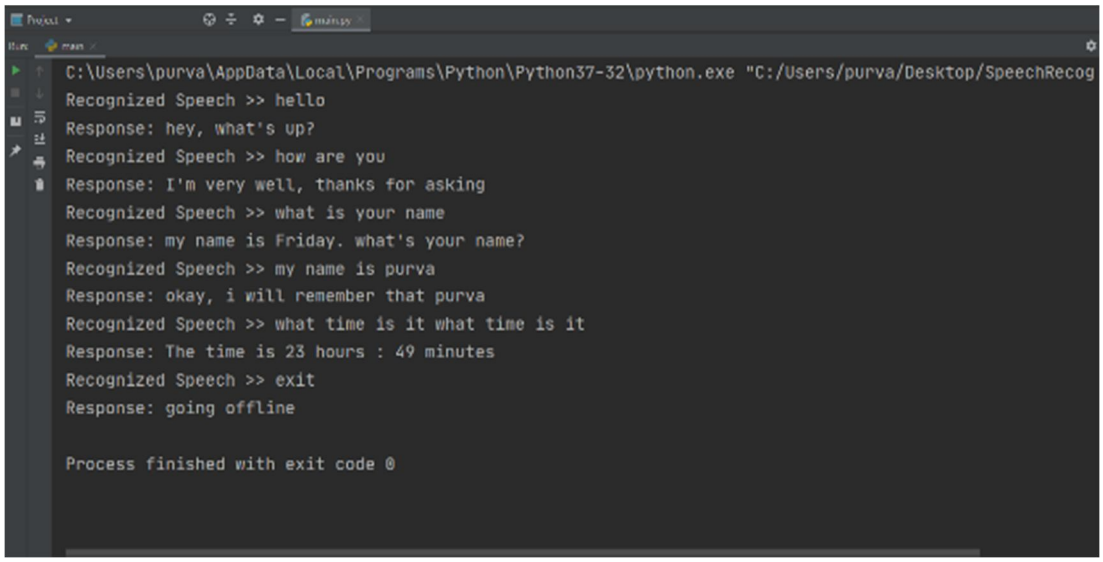


Figure 5: Recognized Speech and Response

As show in the above result, the proposed system can act as a personal assistant. Due to the DNN-HMM and DBN it can work even in offline mode.

©IJRASET: All Rights are Reserved

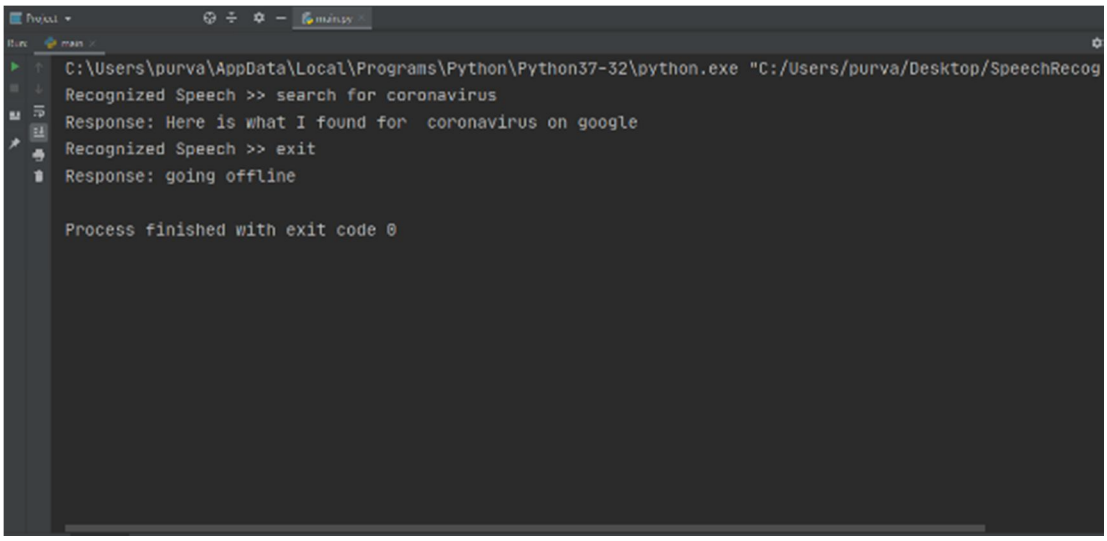


Figure 6: Performing Google Search

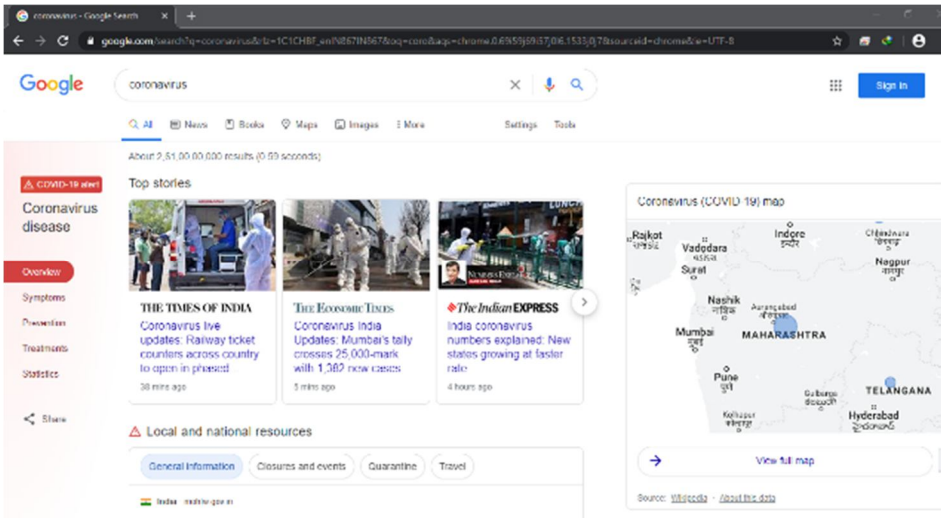


Figure 7: Response of Google

After giving command to the system, it uses the pre trained DBN to perform recognition and performs the actions. As tested above it can search on Google when provided with a search term.

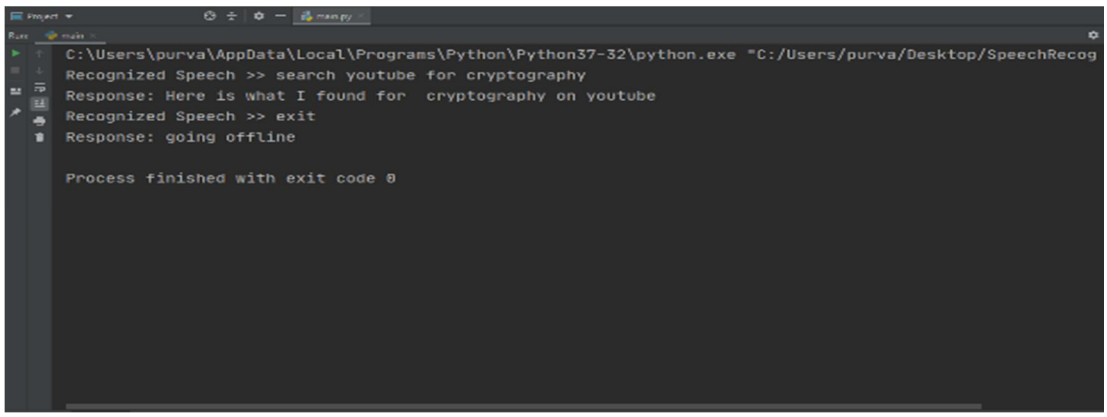


Figure 8: Performing YouTube Search

©IJRASET: All Rights are Reserved

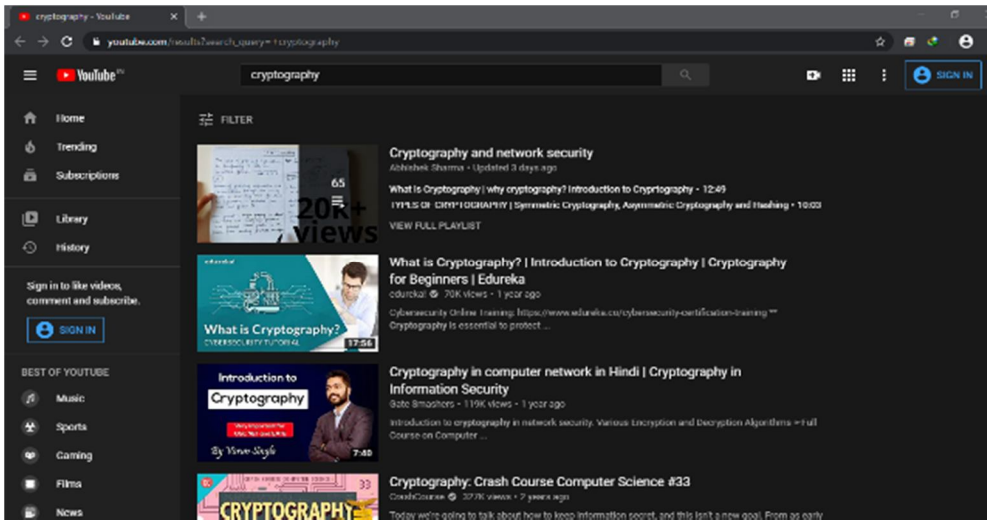


Figure 9: Response of YouTube Search

The above action of YouTube search is performed in less than 0.05 sec which is comparatively very fast than other systems.

VI. CONCLUSION

We developed an Application for continuous speech recognition based on DNN-HMM and DBN algorithms busing python speech features. The projected model is simple to use, low-cost. This model keeps an account and uses of the present developments and numerous sorts of oftenest identification and detection technologies that are used for speech recognition. Due to this once the DBN is trained with sufficient data it can also work in offline mode. The use of DNN-HMM with DBN is giving us the accuracy of 92.34% whereas traditional GMM-HMM system gives 71.8% accuracy which is 20.54% more accurate. The time potency can increase phenomenally since this method can eliminate the stationary process of speech recognition and perform speedy actions accordingly. Automatic speech signals may be served in real time which can be further used in speech translation in real time.

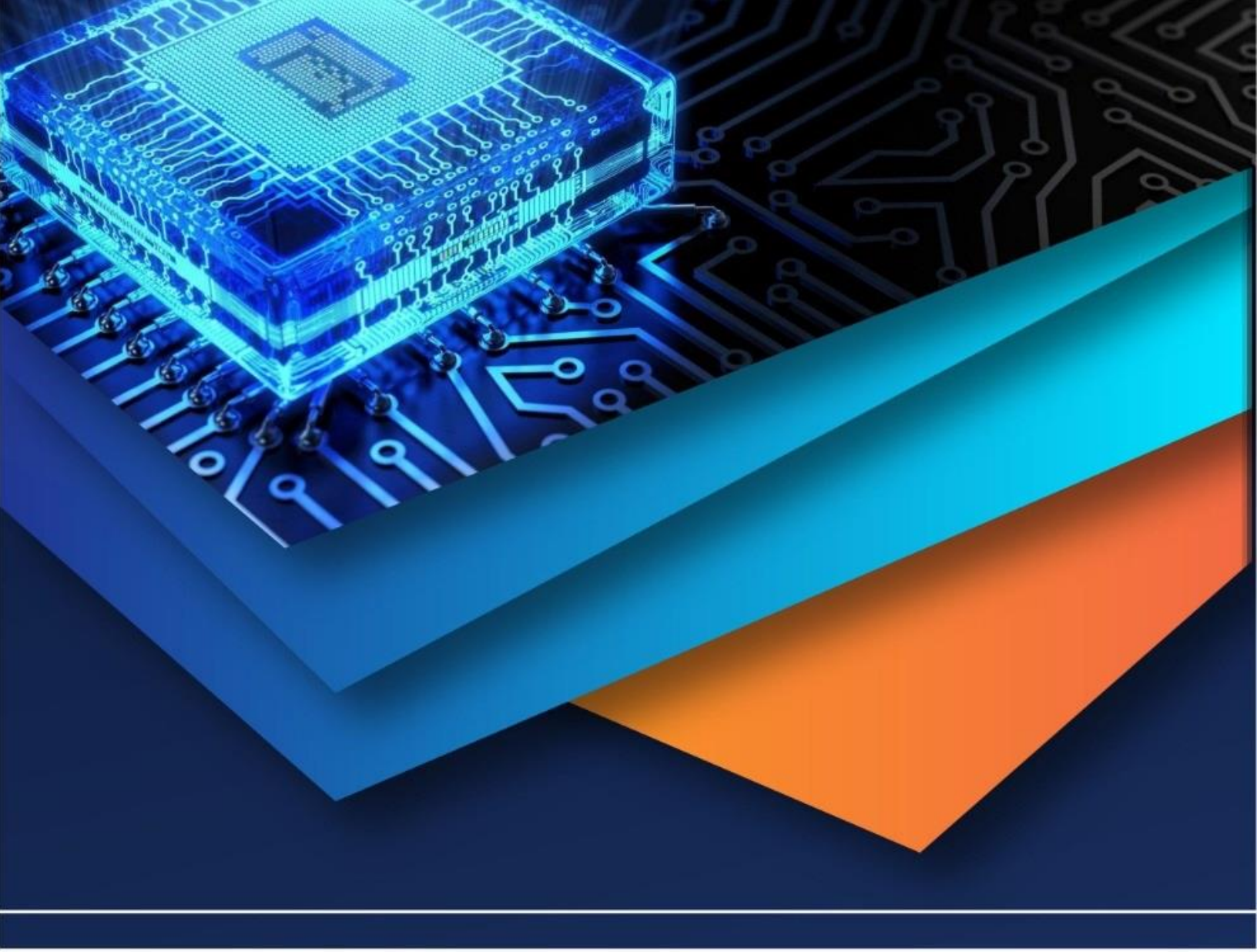
VII. ACKNOWLEDGEMENT

I would like to take the opportunity to express my heartfelt gratitude to the people whose help and co-ordination has made this project a success. I thank Prof. Kirti Rajadnya for knowledge, guidance and co-operation in the process of making this project. I owe project success to my guide and convey my thanks to her. I would like to express my heartfelt to all the teachers and staff members of Information Technology Engineering department of Shivajirao S. Jondhale College for their full support. I would like to thank my principal for conductive environment in the institution. I am grateful to the library staff of Shivajirao S. Jondhale College for the numerous books, magazines made available for handy reference and use of internet facility. Lastly, I am also indebted to all those who have indirectly contributed in making this project successful.

REFERENCES

[1] 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [1] By W. Xiong, L. Wu, F. Alleva, J. Droppo , X. Huang, A. Stolcke. (Microsoft AI and Research.)
[2] A multi- Task Learning Framework for Overcoming the catastrophic forgetting in automatic speech recognition By - Jiabin Xue, Jiqing Han, Tieran Zheng, Xiang Gao, Jiaxing Guo. (Submitted on 17 Apr 2019)
[3] International Journal of Knowledge-based and Intelligent Engineering Systems, (21 March 2018) By Haridas, Arul Valiyavalappil, Marimuthu, Ramalatha, Sivakumar, Vaazi GAngadharan.
[4] Speech Emotion Recognition Using Deep Neural Network Considering Verbal and Nonverbal Speech Sounds By Kun-Yi Huang; Chung-Hsien Wu ; Qian-Bei Hong ; Ming-Hsiang Su ; Yi-Hsuan Chen.
[5] Deep factorization for speech signal (2014) By Lantian Li, Dong Wang, Yixiang Chen, Ying Shi, Zhiyuan Tang, Thomas Fang Zheng.
[6] Leveraging automatic speech recognition in cochlear implants for improved speech intelligibility under reverberation (2018) By Shabnam Ghaffarzadegan, John H.L. Hansen Center for Robust Speech Systems (CRSS).
[7] Listen, Attend and Spell (2015) By William Chan, Navdeep Jaitly, Quoc V. Le, Oriol Vinyals (Google Brain).

©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)