IEEE TENSYMP 2020
Technology for Impactful Sustainable Development
5-7 June 2020, Dhaka, Bangladesh

IEEE region 10

IEEE Bangladesh Section

5-7 June, Dhaka, Bangladesh

Effects of Noise on RASTA-PLP and MFCC based Bangla ASR Using CNN

AUTHOR'S NAME:

MD. RAFFAEL MARUF MD. OMAR FARUQUE

SALMAN MAHMOOD NAZMUN NAHAR NELIMA

MD. GOLAM MUHTASIM MD.JAHEDUL ALAM PARVEZ

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING,

AHSANULLAH UNIVERSITY OF SCIENCE AND TECHNOLOGY,DHAKA

Outline of the Presentation

This short and informative presentation consists of some vital points:

- **Introduction of ASR**
- **Basic description on two feature extraction methods (RASTA-PLP & MFCC)**
- **CNN Model as a classifier and it's architecture**
- **Methodology and Dataset for the specific paper**
- **Result and further discussion**
- **Conclusion and Future works**

INTRODUCTION

What is ASR?

Automatic Speech Recognition or **ASR**, as it's known in short, is the technology that allows human beings to use their voices to speak with a computer interface in a way that, in its most sophisticated variations, resembles normal human conversation. Despite the prospectus of ASR, Bangla, the 7th most spoken language with 250 million speakers around the world lags behind others due to the lower number of conducted researches in this field.

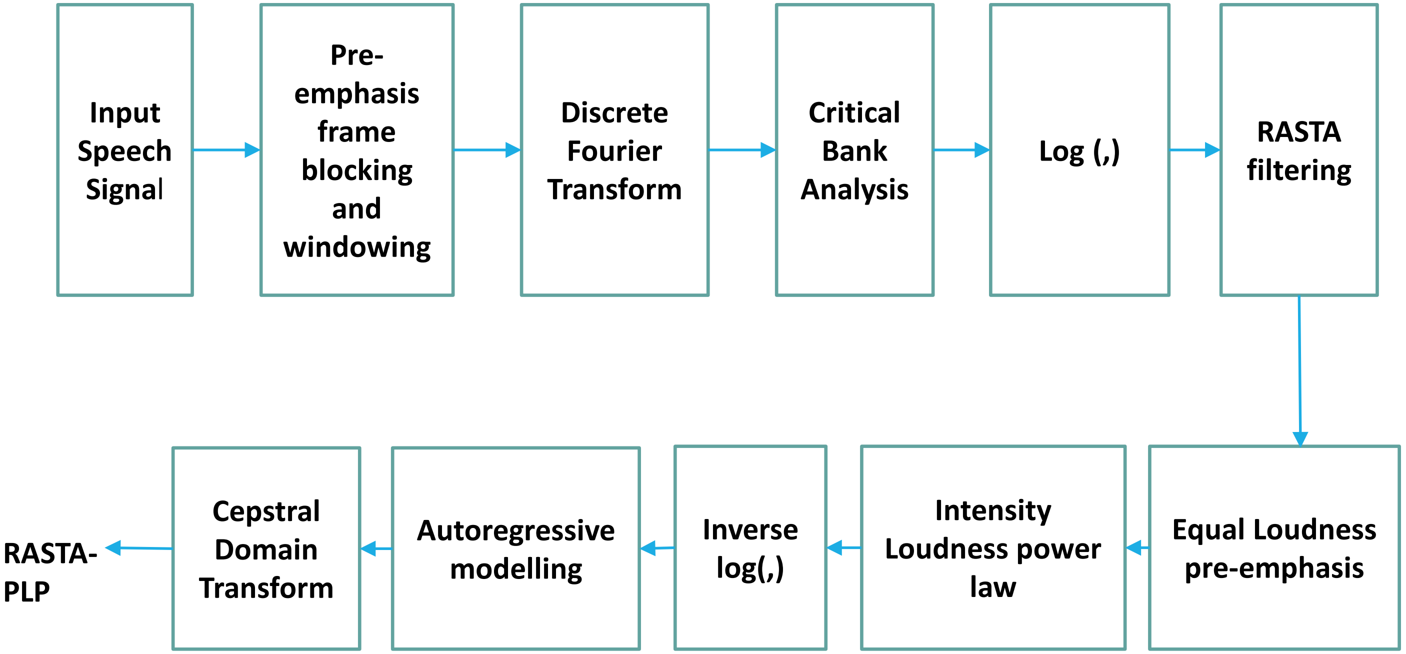
Motivations for this Paper

Though the Bengali literature is so much enriched but very little research works have been done on this language speech recognition. We have set three remarkable applications which can be footprint for the next researchers. They are:

- **Effects of noise in Bangla command words & digits with CNN classifier**
- **The comparative performance of RASTA-PLP and MFCC**
- **Achieving state of the art accuracy in CNN based Bangla ASR which is 93.18%**

Feature Extraction Method

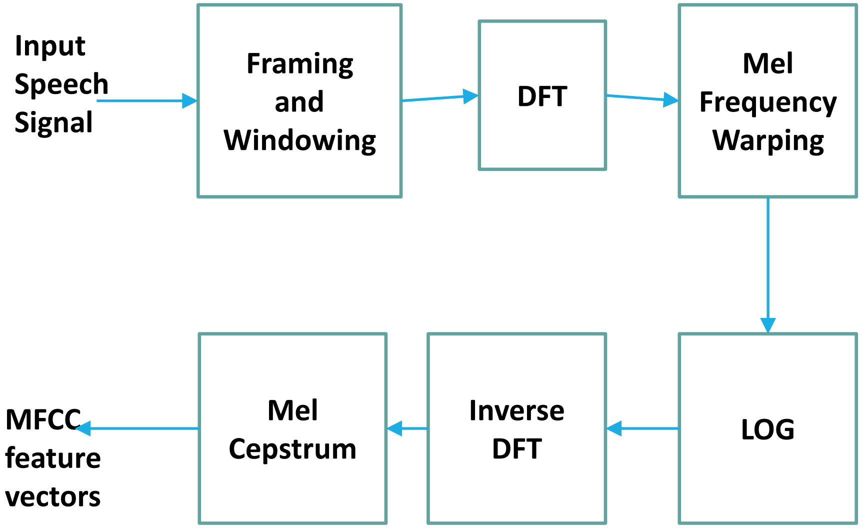
RASTA-PLP



Input Speech Signal → Pre-emphasis frame blocking and windowing → Discrete Fourier Transform → Critical Bank Analysis → Log (,) → RASTA filtering → Equal Loudness pre-emphasis → Intensity Loudness power law → Inverse log(,) → Autoregressive modelling → Cepstral Domain Transform → RASTA-PLP

Block Diagram of RASTA PLP

MFCC



Input Speech Signal → Framing and Windowing → DFT → Mel Frequency Warping → LOG → Inverse DFT → Mel Cepstrum → MFCC feature vectors

Block Diagram of MFCC

Convolutional Neural Network

A convolutional neural network (CNN) is a specific type of artificial neural network that uses perceptrons, a machine learning unit algorithm, for supervised learning, to analyze data. It has five layers :

- □ **Convolutional Layer**
- □ **Pooling Layer**
- □ **Rectified Linear Units (ReLU)**
- □ **Dropout Layer**
- □ **Fully Connected Layer**

Methodology

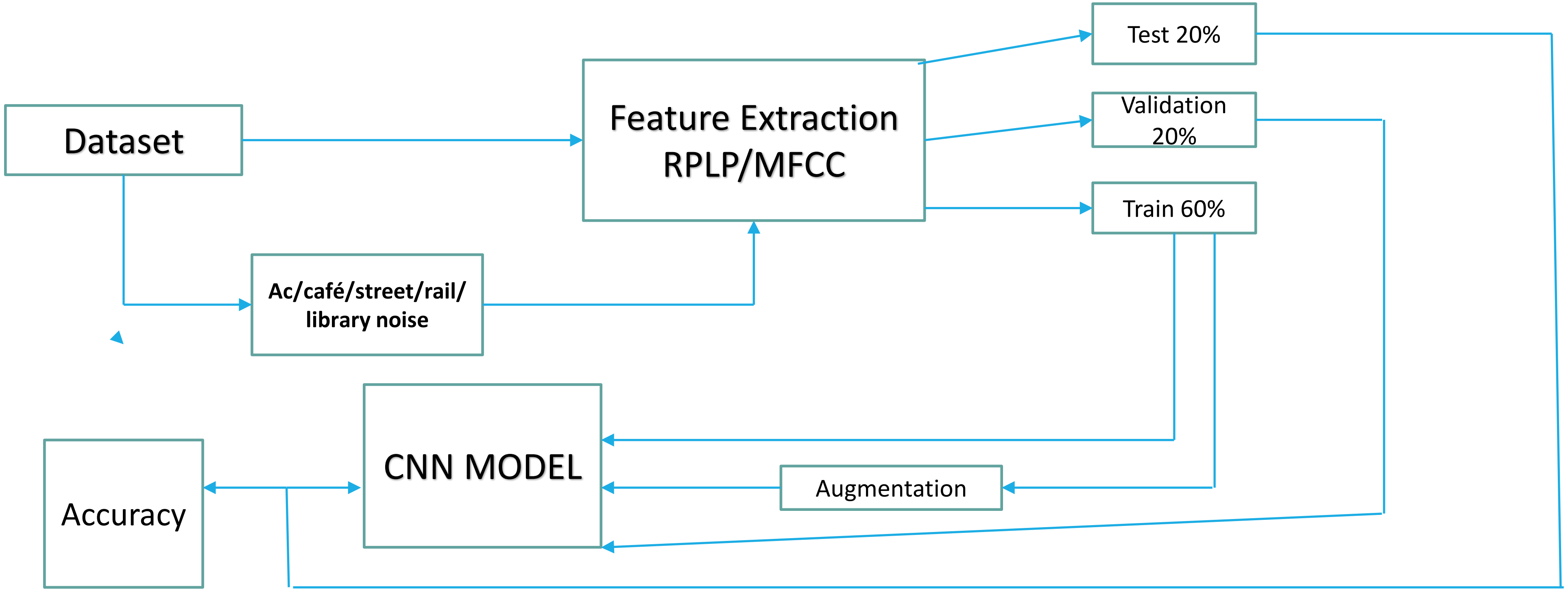


Diagram : Proposed Methodology system of this paper

Dataset

Bengali words	English translation	Phonetic representation
এক	One	Ek
দুই	Two	Dui
তিন	Three	Tin
চার	Four	Char
পাঁচ	Five	Pach
শুরু	Start	Shuru
শেষ	End	Shesh
আসো	Come	Asho
যাও	Go	Jao
ডানে	Right	Dane
বামে	Left	Bame

Table 1: Datasets consists of Isolated Bangla speech commands and digits

CNN Architecture in this paper

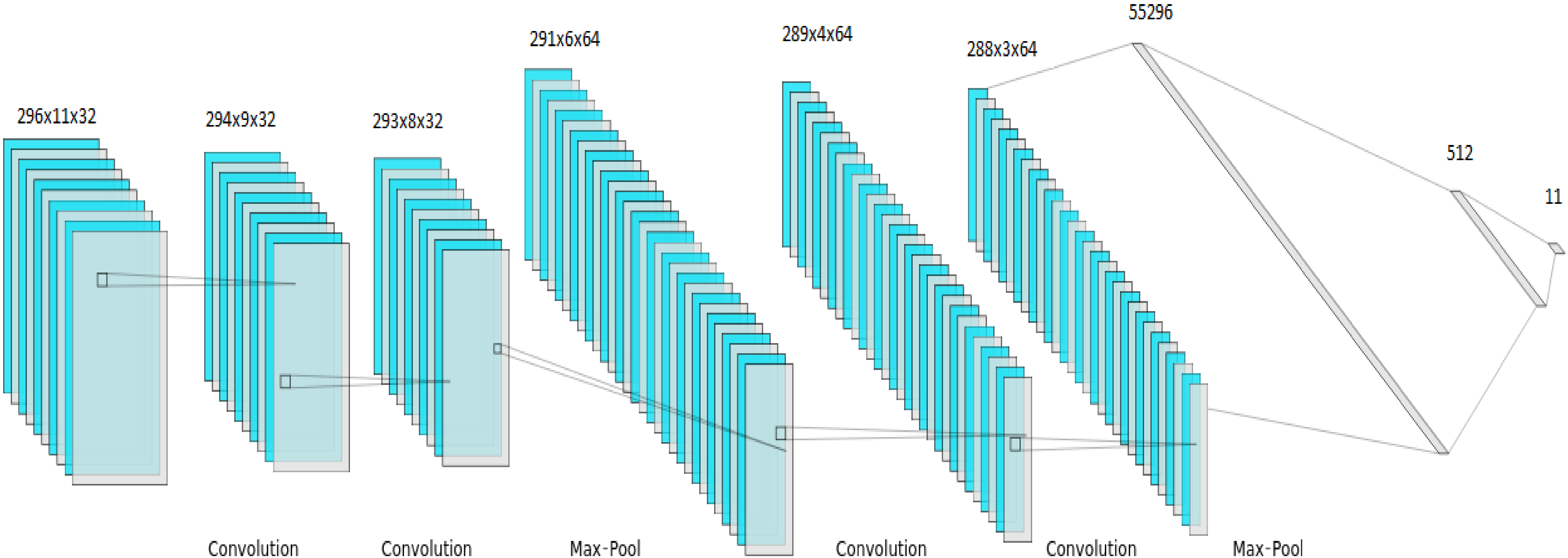


Fig 1: Visualization of CNN Architecture

Results

- The CNN was trained using both MFCC features and RASTA-PLP features extracted directly from without and with augmentation datasets.
- There is a substantial increase in accuracy after augmentation .
- In room environment medium, RASTA-PLP performed better than MFCC model. But in noisy environment, MFCC performed better.
- Overfitting issue is more prominent in RASTA-PLP model than MFCC model.
- Our best accuracy comes when the data is augmented.

Medium	Before augmentation				After augmentation			
	RASTA-PLP		MFCC		RASTA-PLP		MFCC	
	Train	Test	Train	Test	Train	Test	Train	Test
Room Environment	99.56	81.81	99.28	82.72	99.29	93.18	98.78	91.28
Ac noise	99.62	82.42	98.23	78.78	97.85	85.60	97.77	89.01
Cafe noise	98.55	81.51	98.36	78.78	98.82	87.87	98.28	90.15
Library noise	98.3	77.57	97.22	73.03	98.04	85.22	98.40	87.12
Rail station noise	97.6	77.27	87.12	66.66	96.69	82.19	96.10	90.53
Street noise	99.24	80.3	87.12	66.66	98.45	83.33	98.06	87.87

Table 2: Comparison of accuracy between MFCC and RASTA-PLP

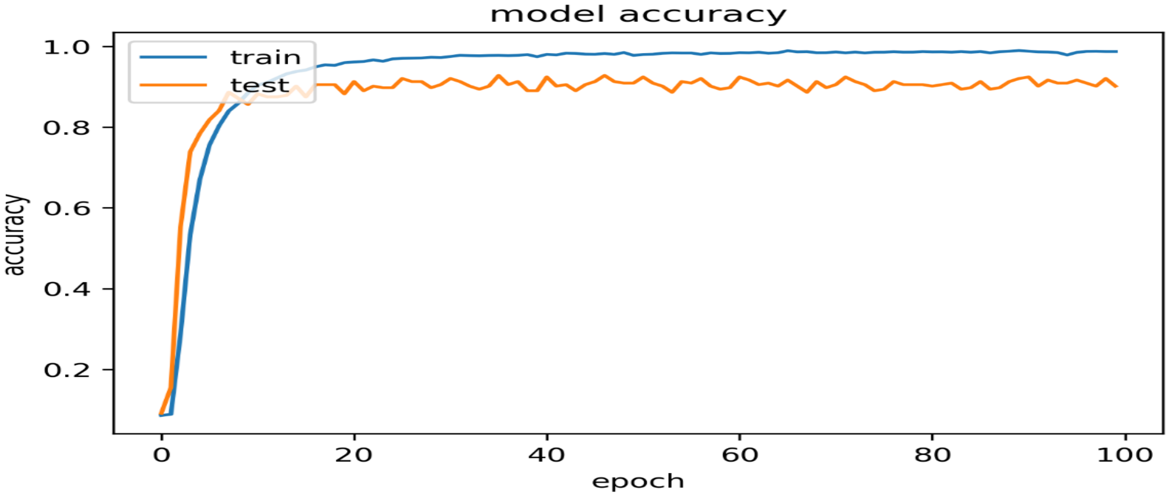


Fig: 2- Epoch vs Accuracy for MFCC in room environment

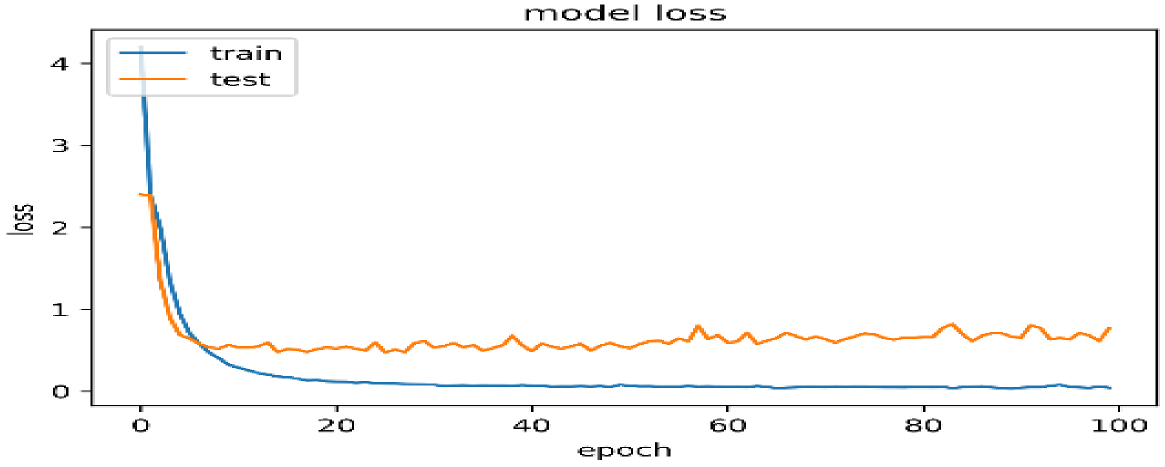


Fig: 3-Epoch vs. Loss for MFCC in room environment.

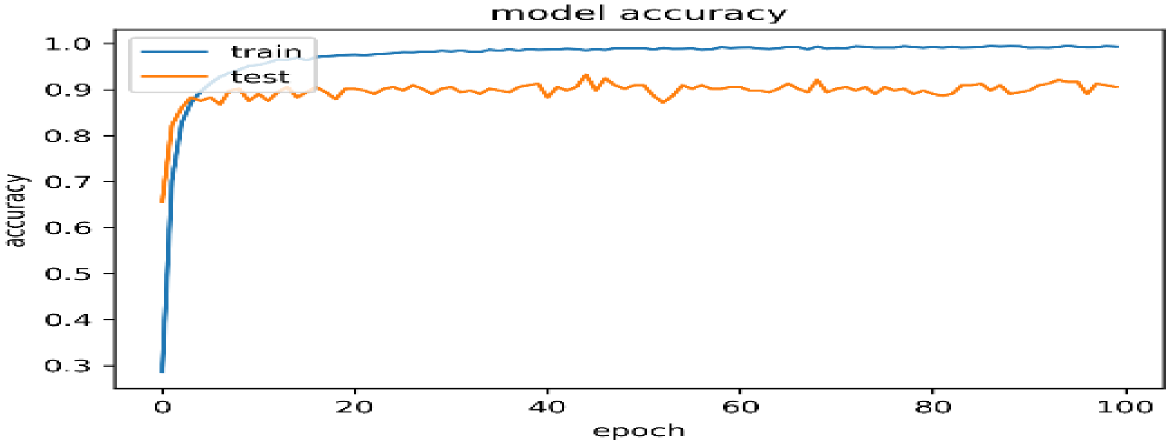


Fig:4-Epoch vs. Accuracy for RASTA-PLP in room environment

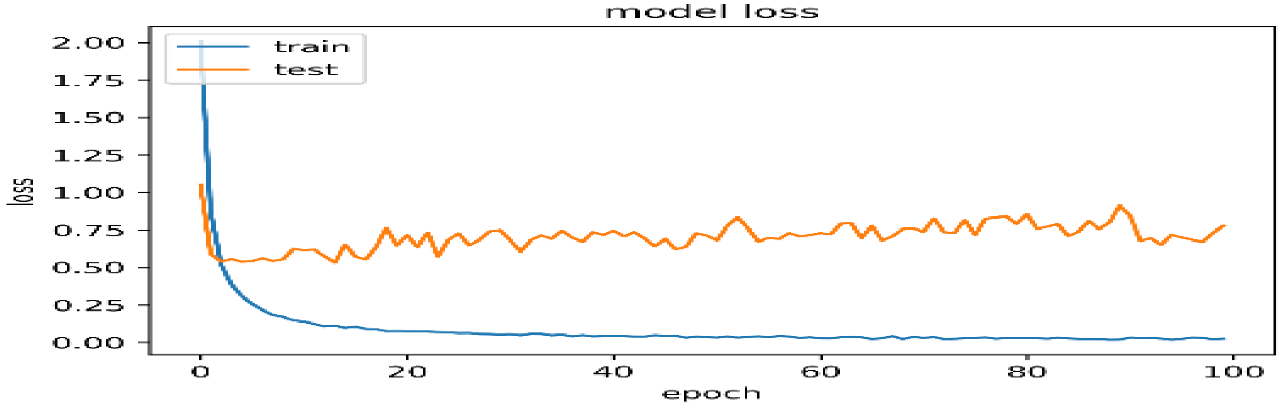
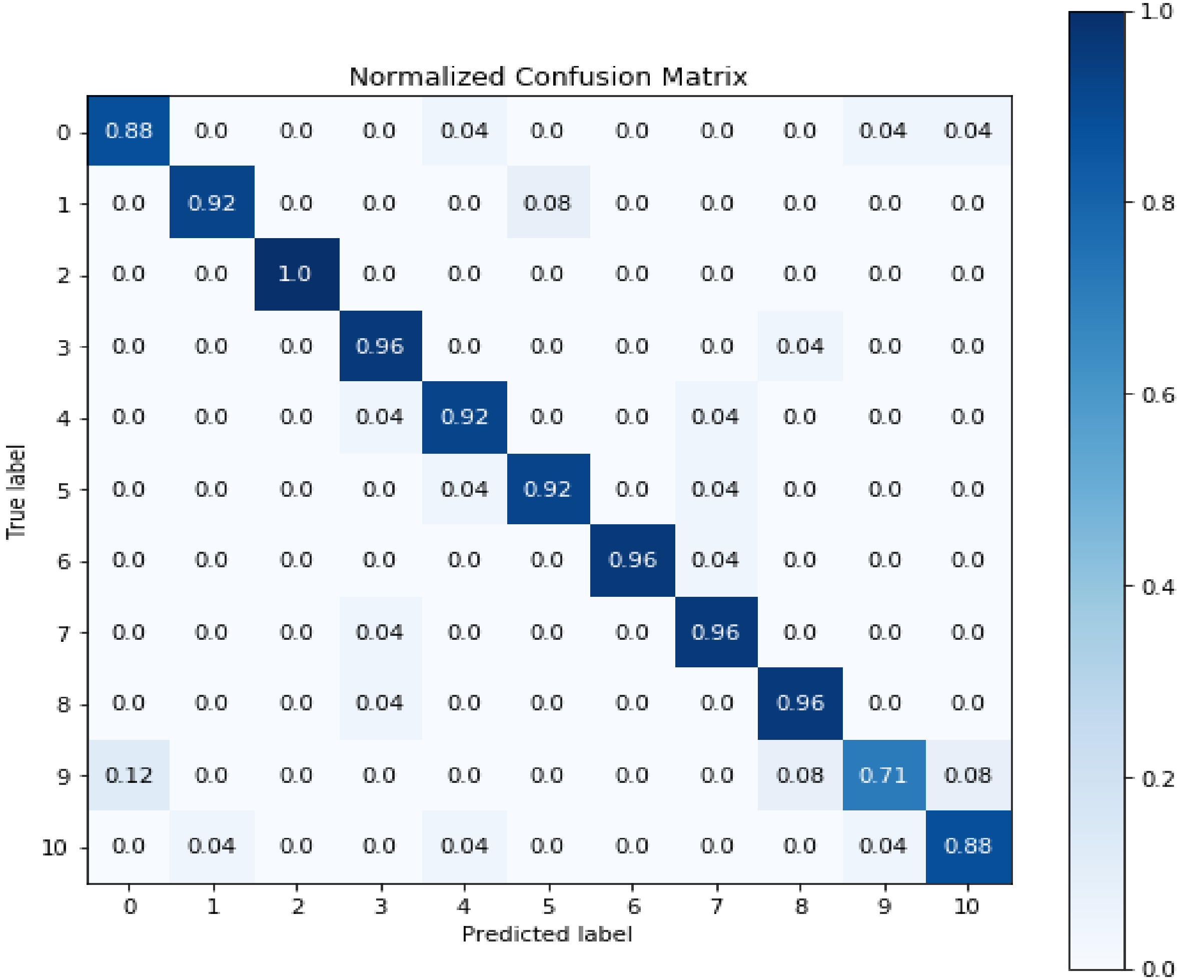
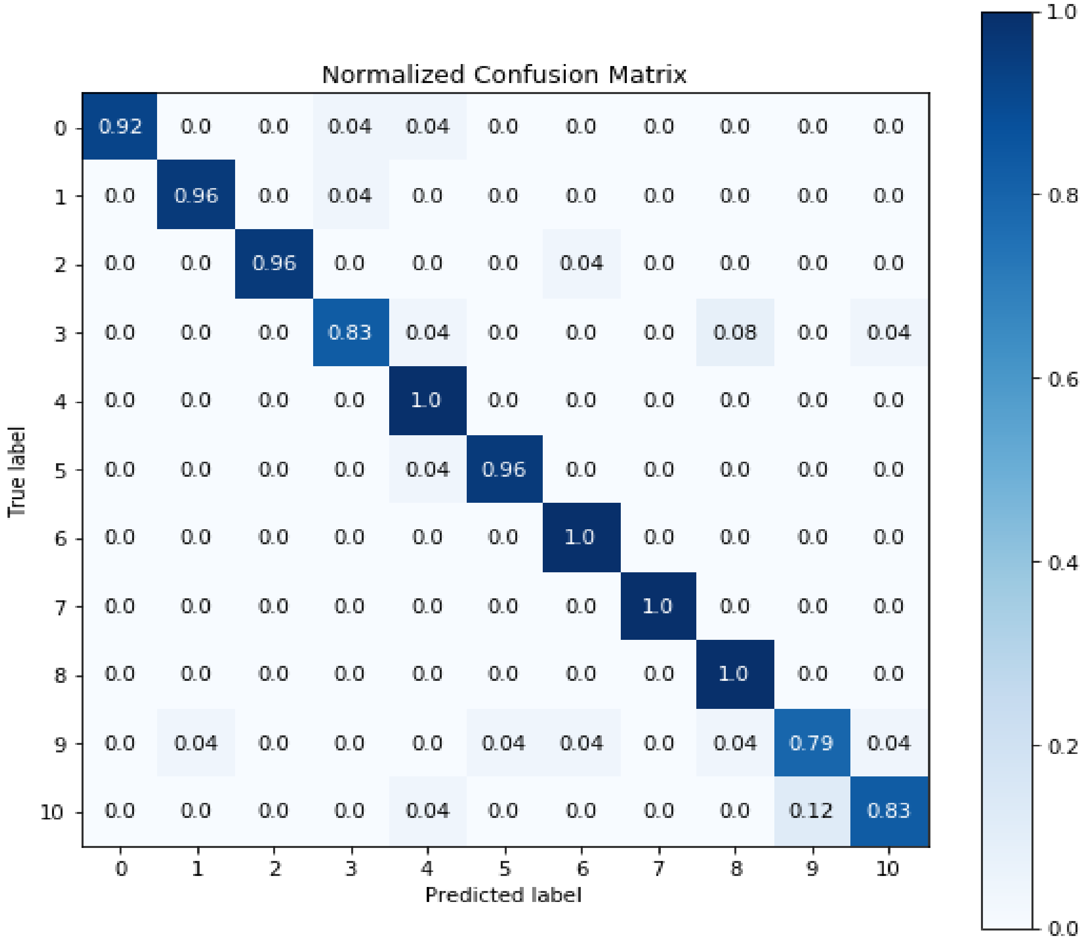


Fig: 5-Epoch vs. Loss for RASTA-PLP in room environment



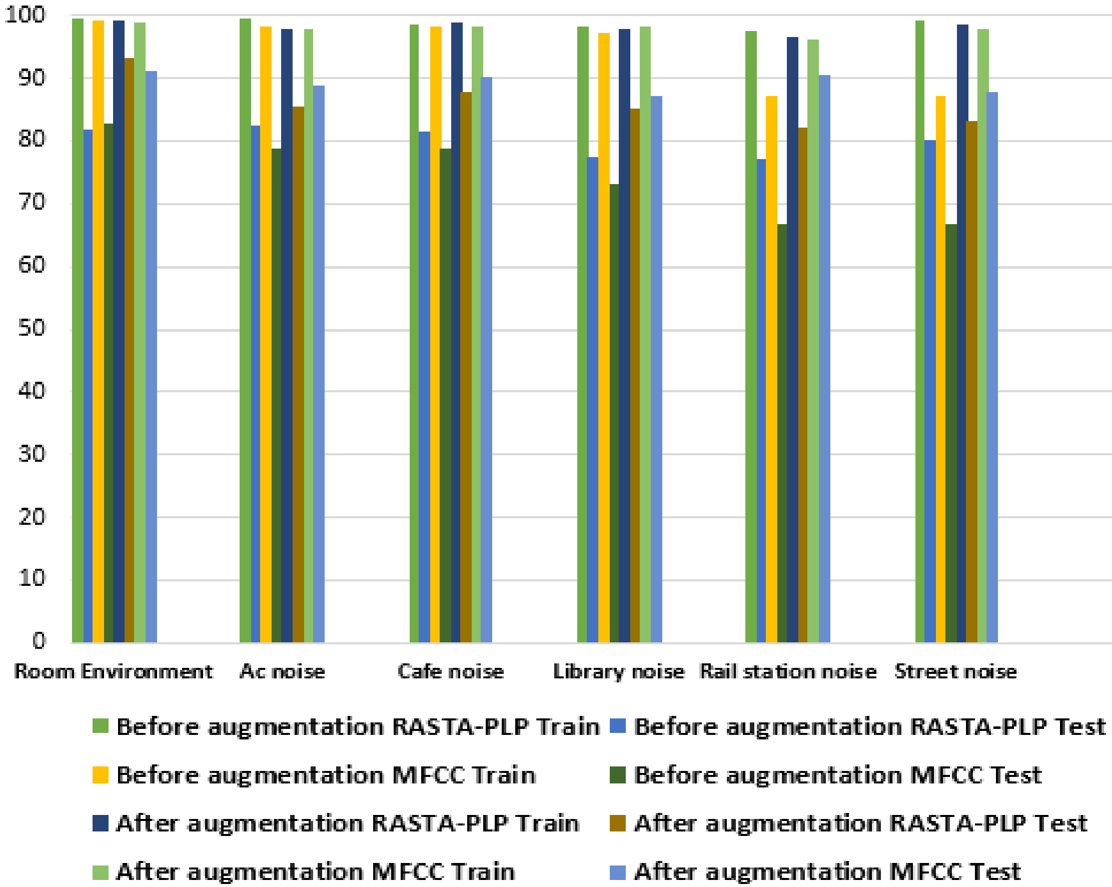
	0	1	2	3	4	5	6	7	8	9	10
0	0.88	0.0	0.0	0.0	0.04	0.0	0.0	0.0	0.0	0.04	0.04
1	0.0	0.92	0.0	0.0	0.0	0.08	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.96	0.0	0.0	0.0	0.0	0.04	0.0	0.0
4	0.0	0.0	0.0	0.04	0.92	0.0	0.0	0.04	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.04	0.92	0.0	0.04	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.96	0.04	0.0	0.0	0.0
7	0.0	0.0	0.0	0.04	0.0	0.0	0.0	0.96	0.0	0.0	0.0
8	0.0	0.0	0.0	0.04	0.0	0.0	0.0	0.0	0.96	0.0	0.0
9	0.12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.08	0.71	0.08
10	0.0	0.04	0.0	0.0	0.04	0.0	0.0	0.0	0.0	0.04	0.88

Fig: 6-Confusion Matrix of MFCC model in room environment.



	0	1	2	3	4	5	6	7	8	9	10
0	0.92	0.0	0.0	0.04	0.04	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.96	0.0	0.04	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.96	0.0	0.0	0.0	0.04	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.83	0.04	0.0	0.0	0.0	0.08	0.0	0.04
4	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.04	0.96	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
9	0.0	0.04	0.0	0.0	0.0	0.04	0.04	0.0	0.04	0.79	0.04
10	0.0	0.0	0.0	0.0	0.04	0.0	0.0	0.0	0.0	0.12	0.83

Fig: 7-Confusion Matrix of RASTA-PLP model in room environment.



Fig;8-Comparison of model accuracy between MFCC and RASTA-PLP.

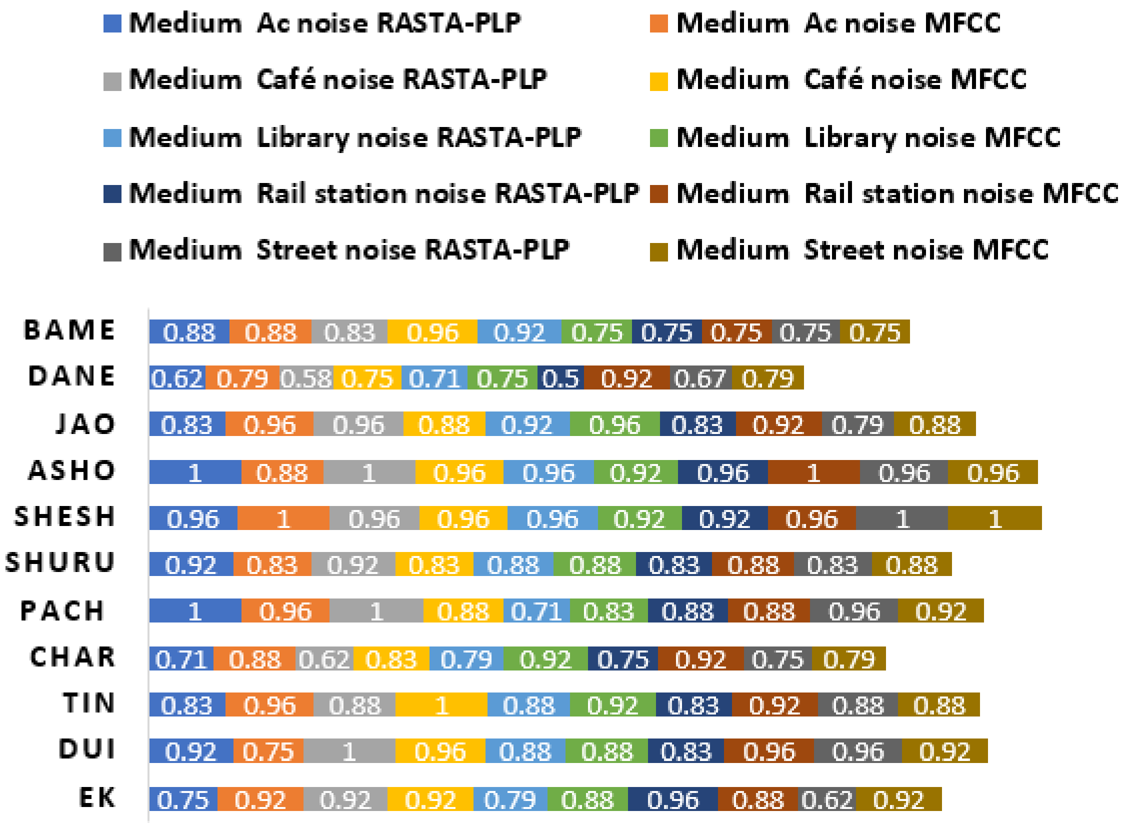


Fig: 9-Comparison of prediction accuracy of various words between RASTA-PLP and MFCC models on various mediums.

TABLE 3– COMPARISON OF ACCURACY BETWEEN PROPOSED CNN BASED MODEL WITH CONTEMPORARY CNN BASED MODEL:

Model	Train accuracy	Test accuracy
Proposed RASTA-PLP model (Room Environment)	99.29%	93.18%
Reference MFCC model	85.44%	74.01%

Conclusion

- . Experimental results show that RASTA-PLP outperforms MFCC in room environment but MFCC performs better in noisy environment.
- In the future, we would like to increase the vocabulary range to implement ASR model in hardware system
- Bangla voice-controlled wheelchairs as well as robotic arm can be an example.