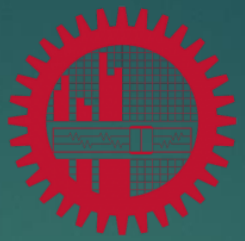


BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY

CONTINUOUS SPEECH RECOGNITION

COURSE NO. : EEE 312

COURSE TITLE : DIGITAL SIGNAL PROCESSING I LABORATORY

PRESENTED BY:

STUDENT ID: 2006010,2006031,
2006032,2006033

Group :06
Section :A1
Level-3,Term-1

Outline

1. Objectives
2. Abstract
3. Workflow
4. Collecting Voice Samples
5. Data preprocessing and augmentation
6. Generating train data set and validation data set
7. Extracting features
8. Building CNN model
9. Word by word segmentation
10. Speech detection by trained model
11. Program Outcomes

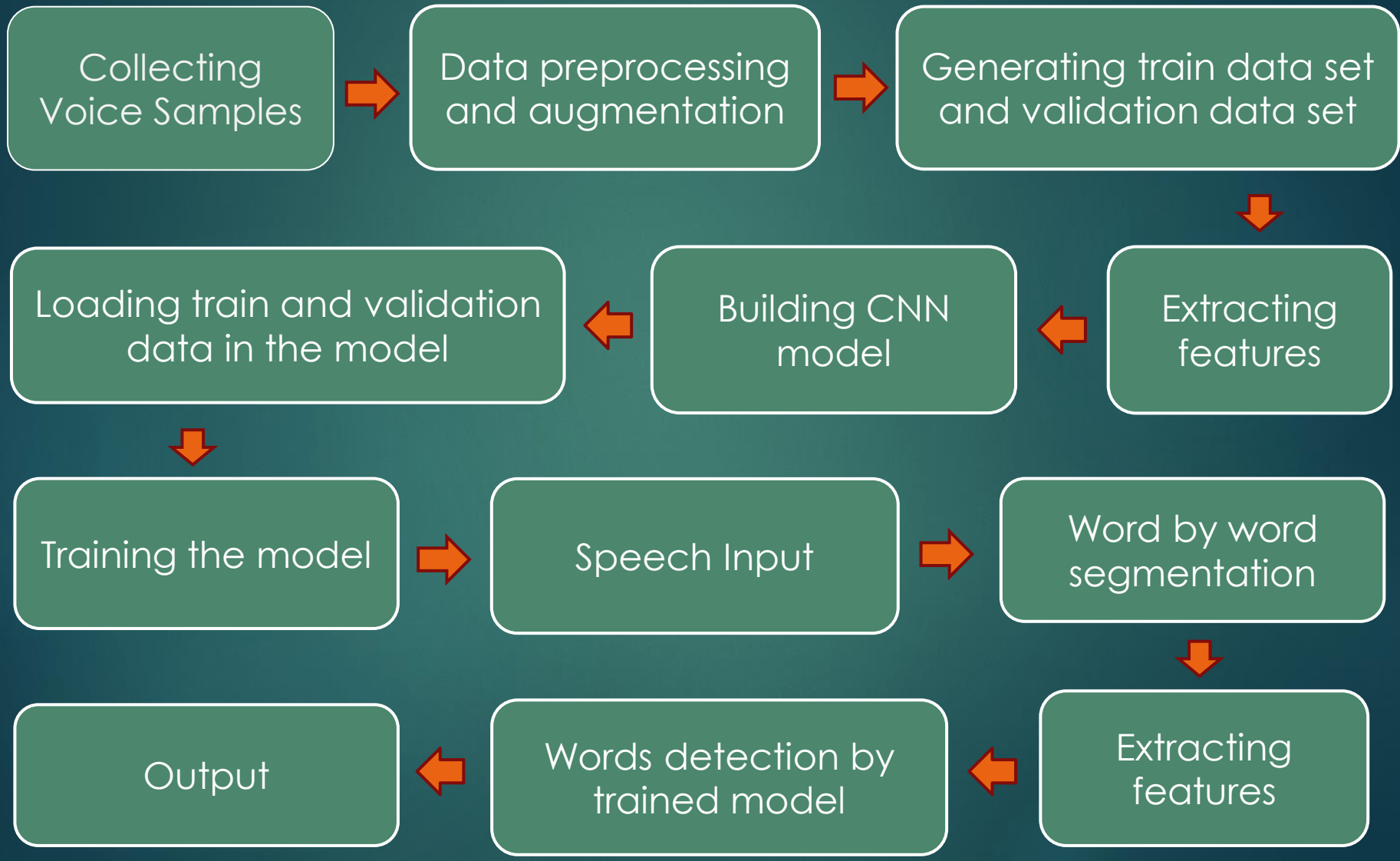
Objectives

- Speech segmentation.
- Feature extraction
- CNN model build up.
- Speech Recognition

Abstract

- Recognizing speech is challenging, especially in noisy real-life conditions.
- Convolutional neural network model is capable to identify speech.
- This project explores various speech processing objectives, including:
  - Recognition
  - Identification
  - Utilizing signal processing
  - Deep learning techniques
- It aims to enhance human-computer interaction, authentication systems, and assistive technologies, contributing to accessibility and innovation.

Workflow


Collecting Voice Samples
Data preprocessing and augmentation
Generating train data set and validation data set
Extracting features
Building CNN model
Loading train and validation data in the model
Training the model
Speech Input
Word by word segmentation
Extracting features
Words detection by trained model
Output

Collecting Voice Samples

- We have 9 different words.
- Which can generate up to 435,848,049 (435 Million) Sentences
- We took 30 raw samples for each of the word for validation data set.
- Total 270 validation data.
- We took 356 samples for each of the word for training data set
- Total 3,204 training data.
- Total 200 background data.

Data preprocessing and augmentation

- Pitch changing
- Adding noise
- Time stretching
- Filtering
- Time shifting
- Volume change

Generating train data set and validation data set

- **Generating Validation Data**

Normalized 30 samples each from 9 words = **Validation Data**

Total Raw Data = 270

- **Generating Train Data**

Normalized preprocessed and augmented data = **Training Data**

Total Data = 3,204

Extracting features

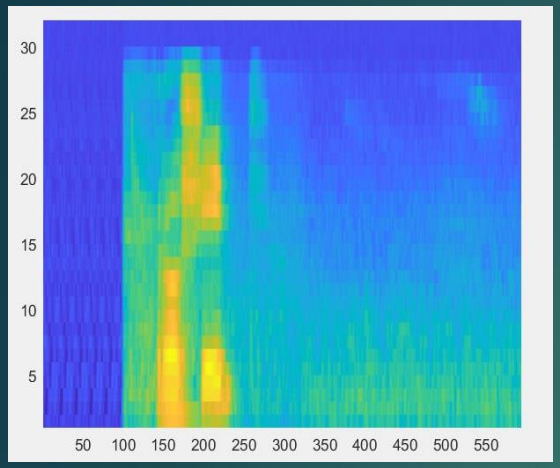
➢ Bark spectrum

The word used are: Bangladesh, Bisshobiddaloy, BUET, Cafeteria, Hall, Nazrul, Prokoushol, Rashid, Sher-e-Bangla.

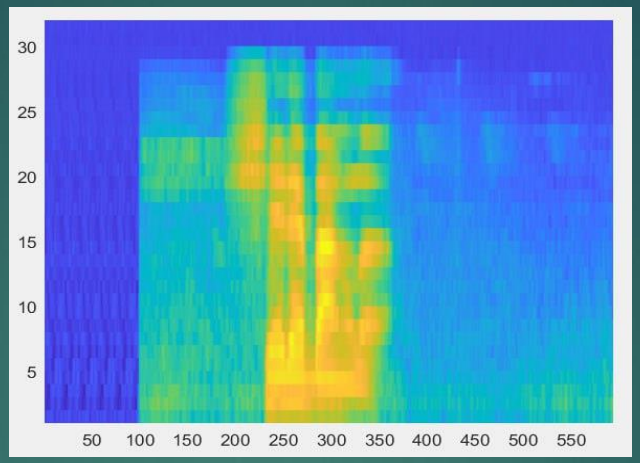
Some example of 5 sentences:

- Bangladesh Prokoushol Bisshobiddaloy
- BUET Cafeteria
- Rashid Hall BUET
- Sher-e-Bangla Hall BUET
- Nazrul Hall BUET

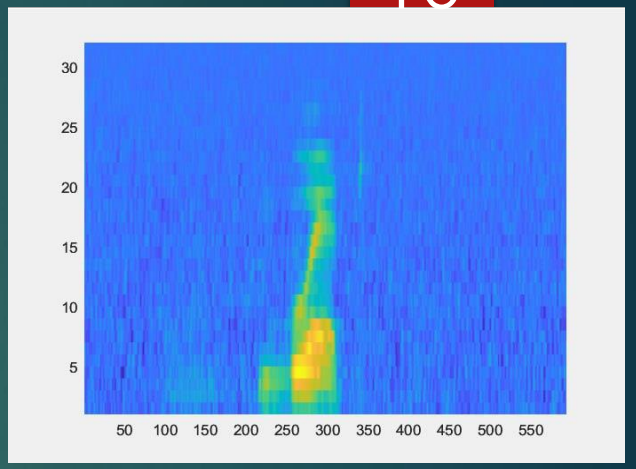
Bark spectrum



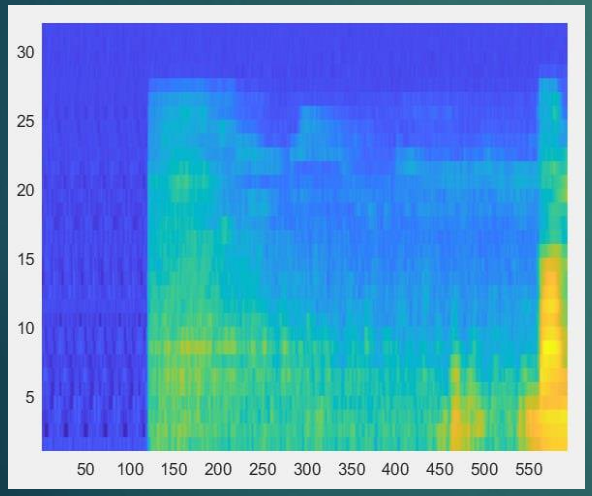
Rashid



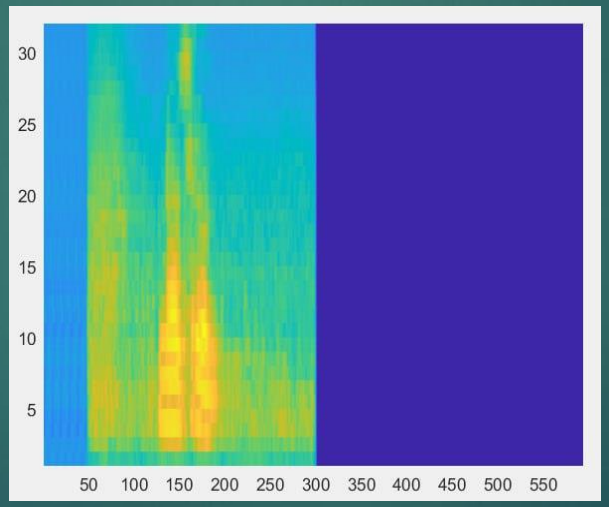
Sher-e-Bangla



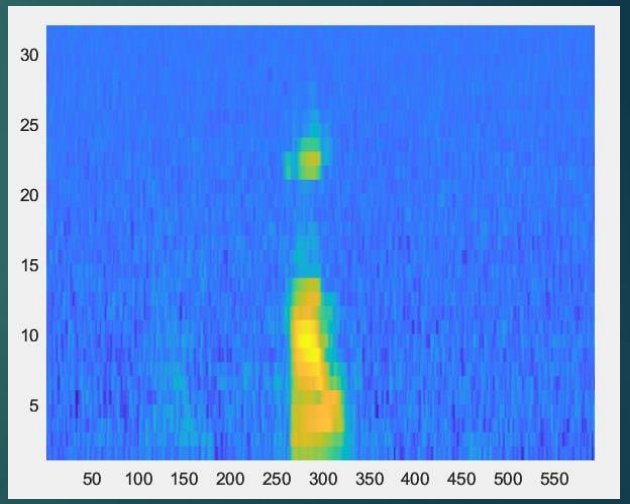
BUET



Bangladesh

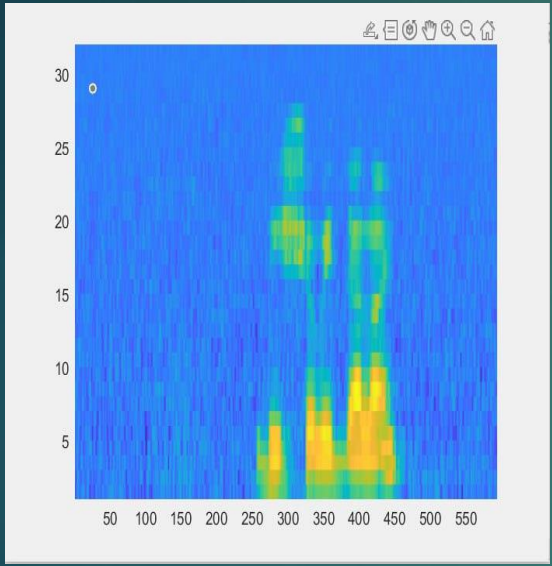


Nazrull

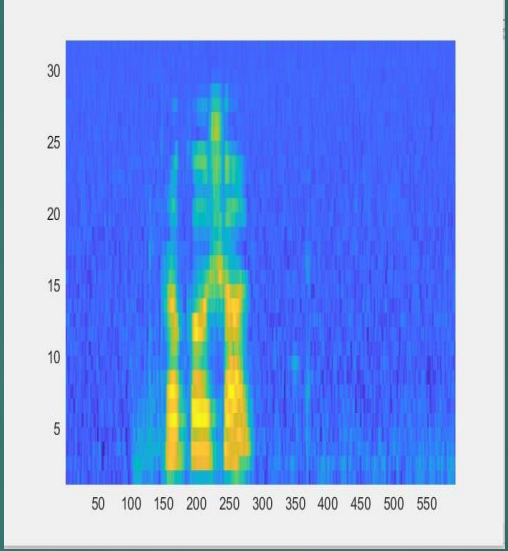


Hall

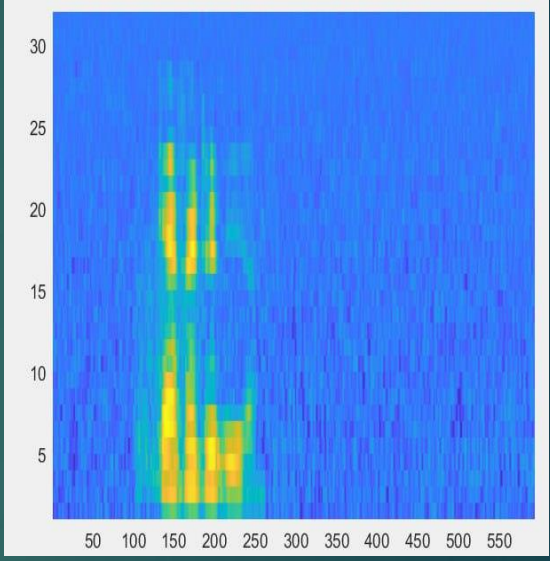
Bark spectrum



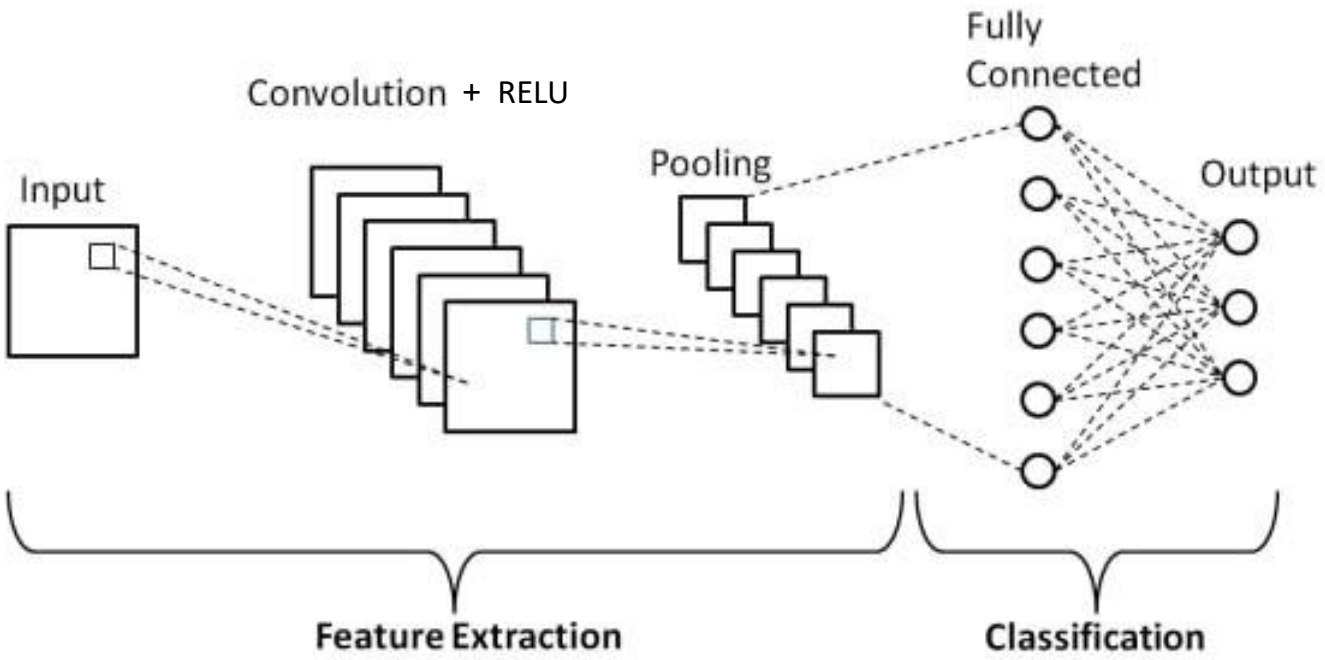
Bisshobiddaloy



Prokoushol

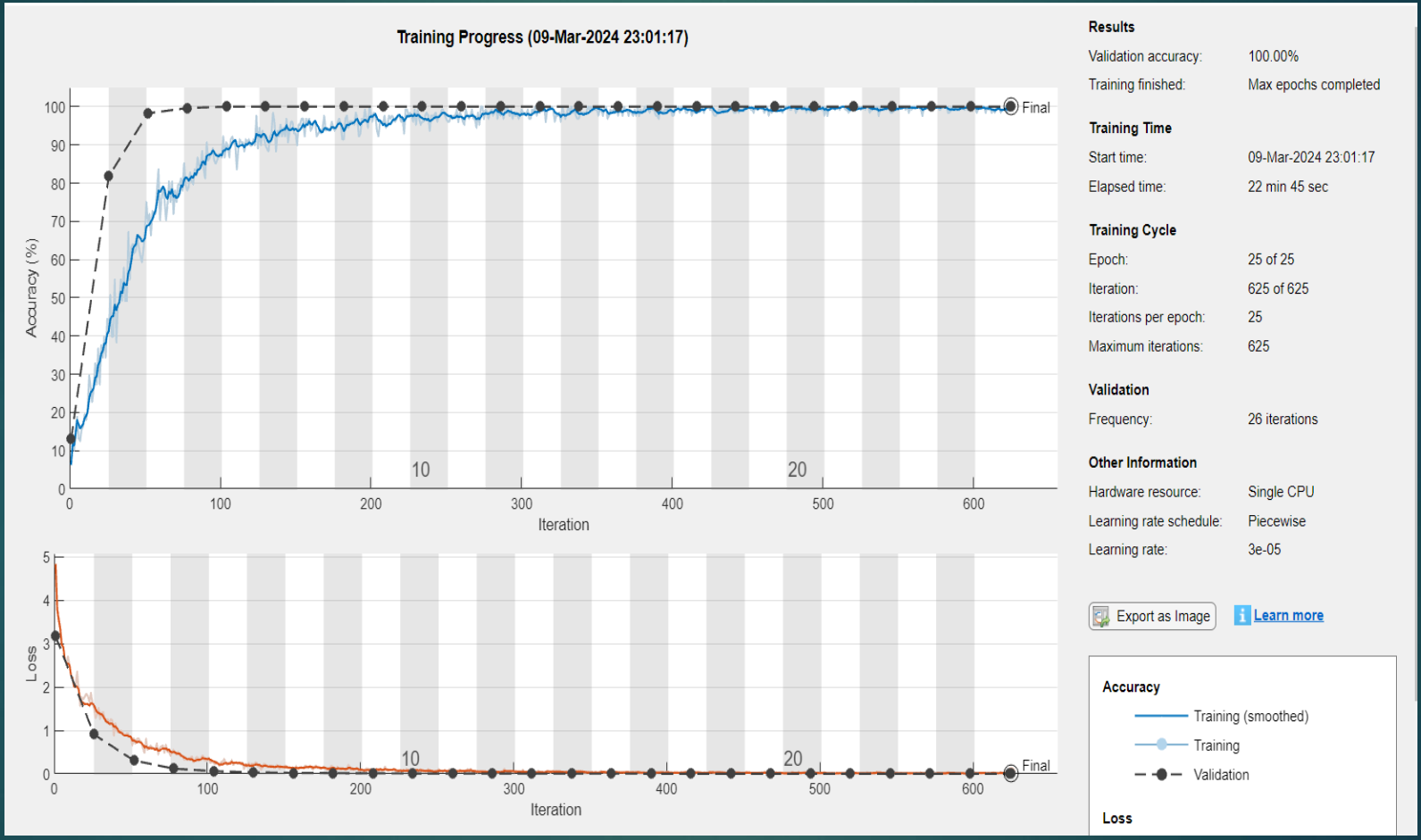


Cafeteria



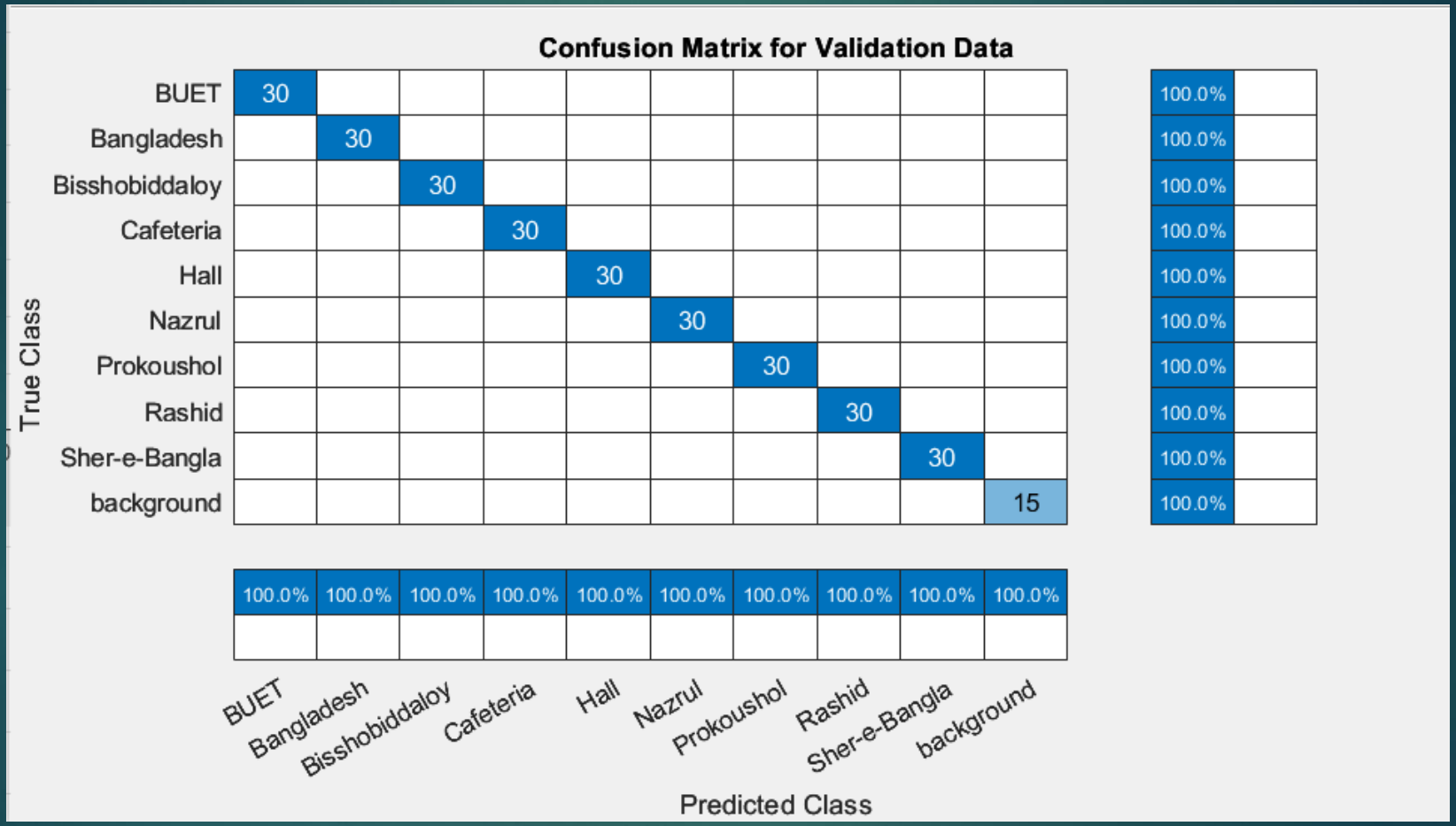
Building CNN model

Training the model



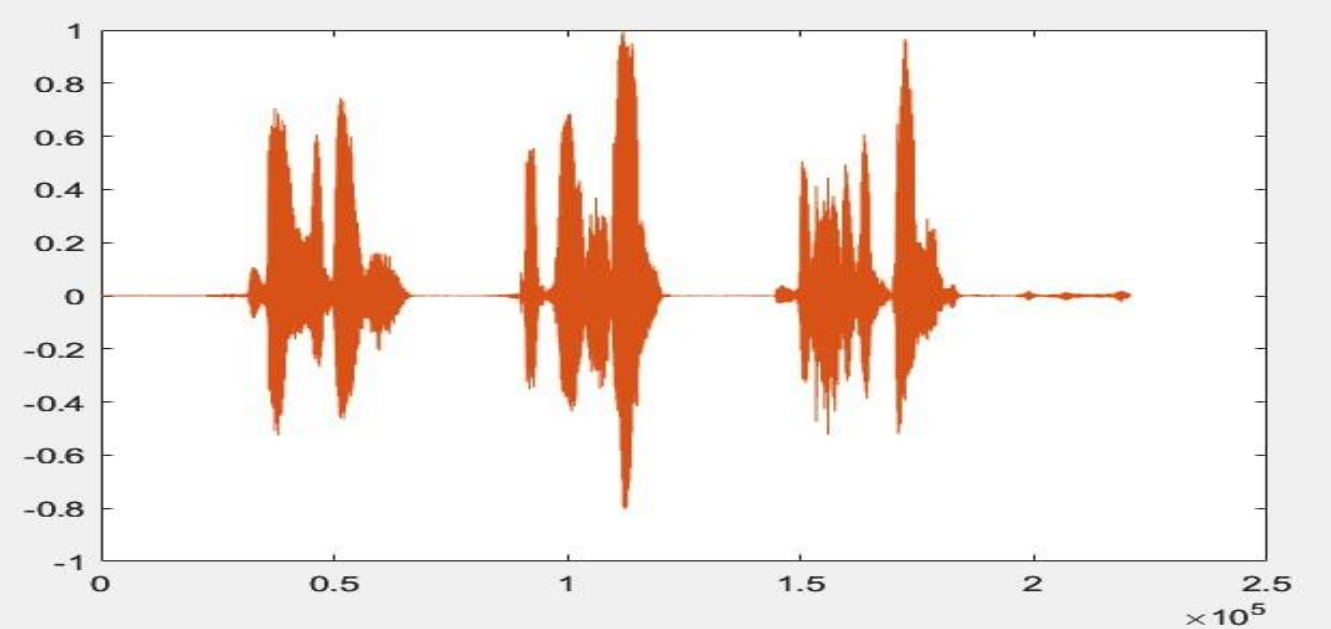
Training Process

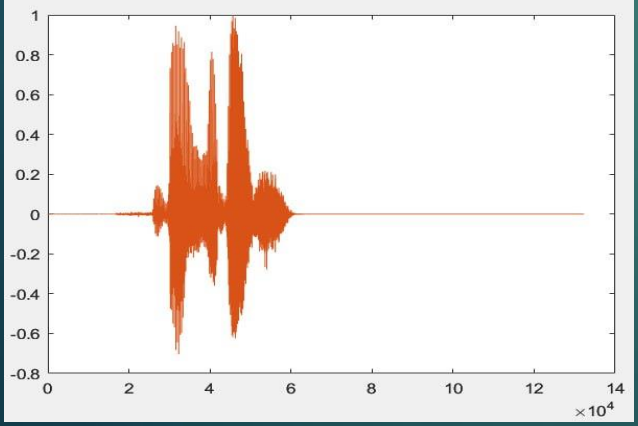
Training the model

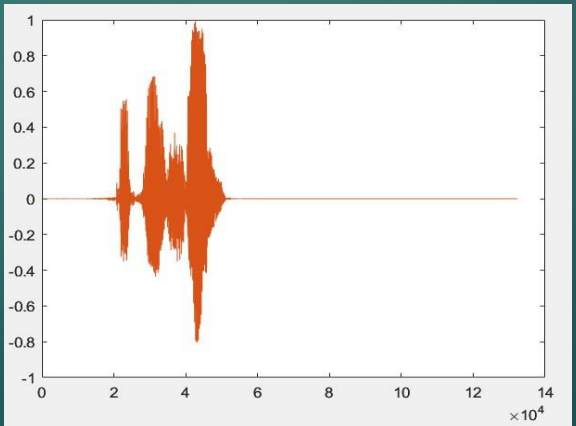


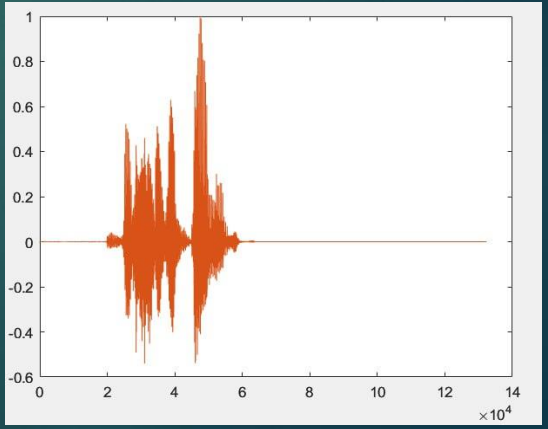
Confusion Matrix for Validation Data

Word by word segmentation

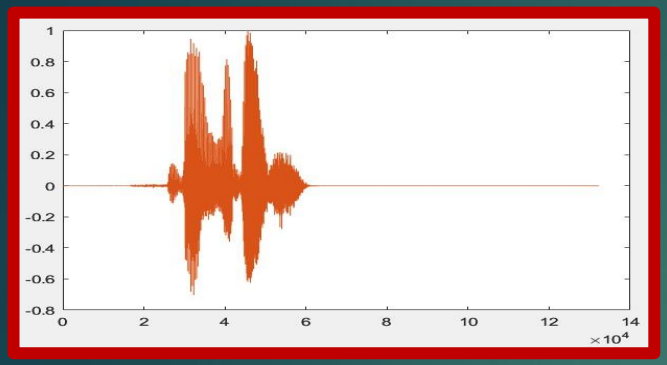




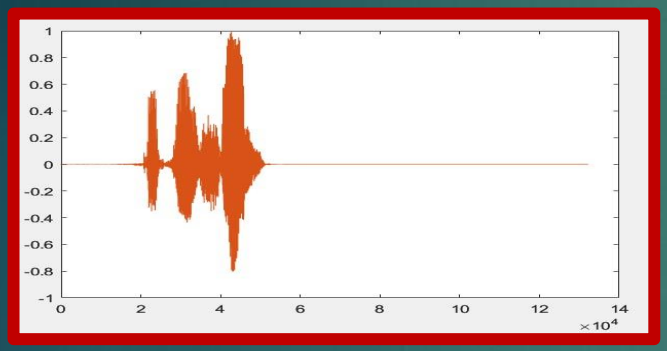




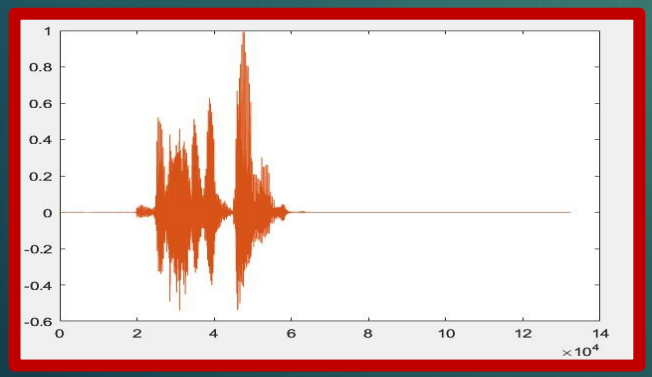
Words detection by trained model



Word -01

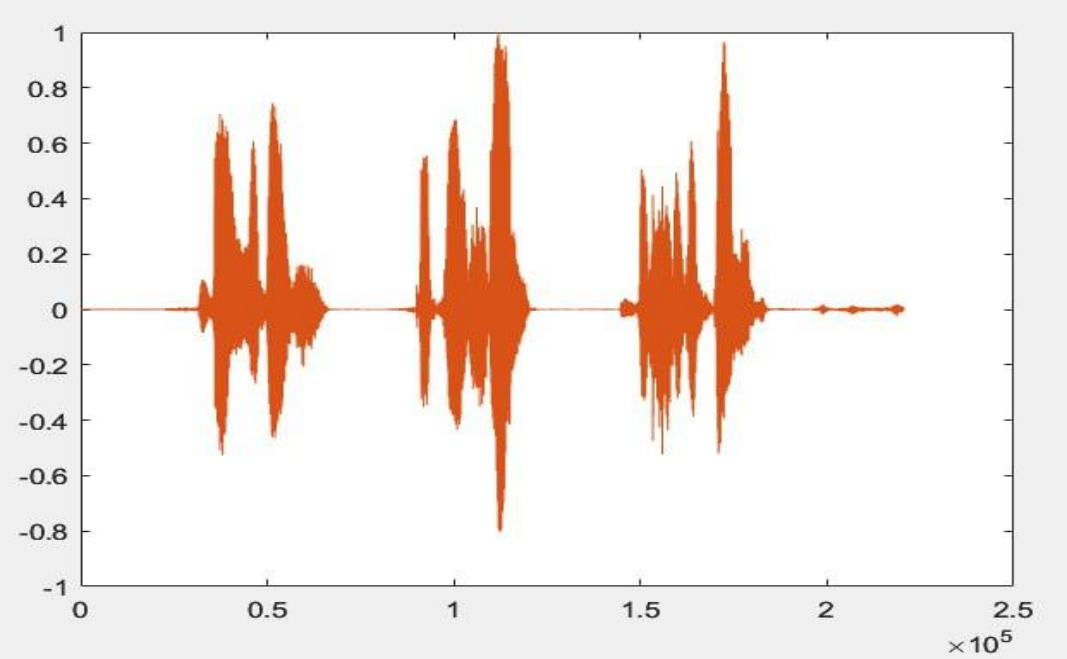


Word -02



Word -03

Trained CNN → Statement



Output

```
>> TestingGround
Start speaking.
End of Recording

Bangladesh
Prokoushol
Bisshobiddaloy
>>
```

PROGRAM OUTCOMES

Program Outcome Achieved In The Project

1.PO(a) Engineering Knowledge:

Knowledge over digital Signal Processing, Convolutional neural network, Audio Feature extrication

2,PO(f) Contextual Knowledge:

Using knowledge that we learn in EEE 311 course

THANK YOU