EEE 312

Digital Signal Processing Laboratory I

Project Title : Bangla voice controlled wheel chair for handicapped persons

Group 1

Submitted to

Dr. Celia Shahnaz
Shahed Ahmed

1706001	Minmoy Kumar Kundu
1706003	Md. Jawad Ul Islam
1706005	Nabila Tasfiha Rahman
1706007	Md. Rafiqul Islam Rafi
1706009	Jahid Hasan Tushar
1706032	Ayan Biswas Pranta

Introduction

Today we are going to present our project on Bangla voice controlled wheel chair for handicapped persons . This will help handicapped people of our country to use their wheelchair without any support from others.

Objectives

1. To control an electric wheelchair with bangla voice command to help physically disabled patients with their day to day life.
2. Create dataset in bangla for commands used in this project
3. To build a CNN model
4. To compare our model with other reference models
5. To create a virtual environment for demonstration

Data Collection

- Our goal was to use bangla voice recognition techniques to detect 5 commands:
    1. Shamne (means forward side)
    2. Pichone (means backward side)
    3. Dane (means right side)
    4. Bame (means left side)
    5. Thamo (means stop moving)
    6. Miscellaneous words for null class
- No open source bangla database suitable for our work, so we created a dataset by taking samples from our classmates, both male and female.
- Around 360 data samples per command, and further augmented, each of length 2 seconds.

Gender ratio :

It is known that there are some characteristics that different in male and female voices for the same speech command.We take this issue in consideration and perform three types of training.

Trained model with only male dataset, only female dataset, both male and female dataset(mixed)

Augmentation:

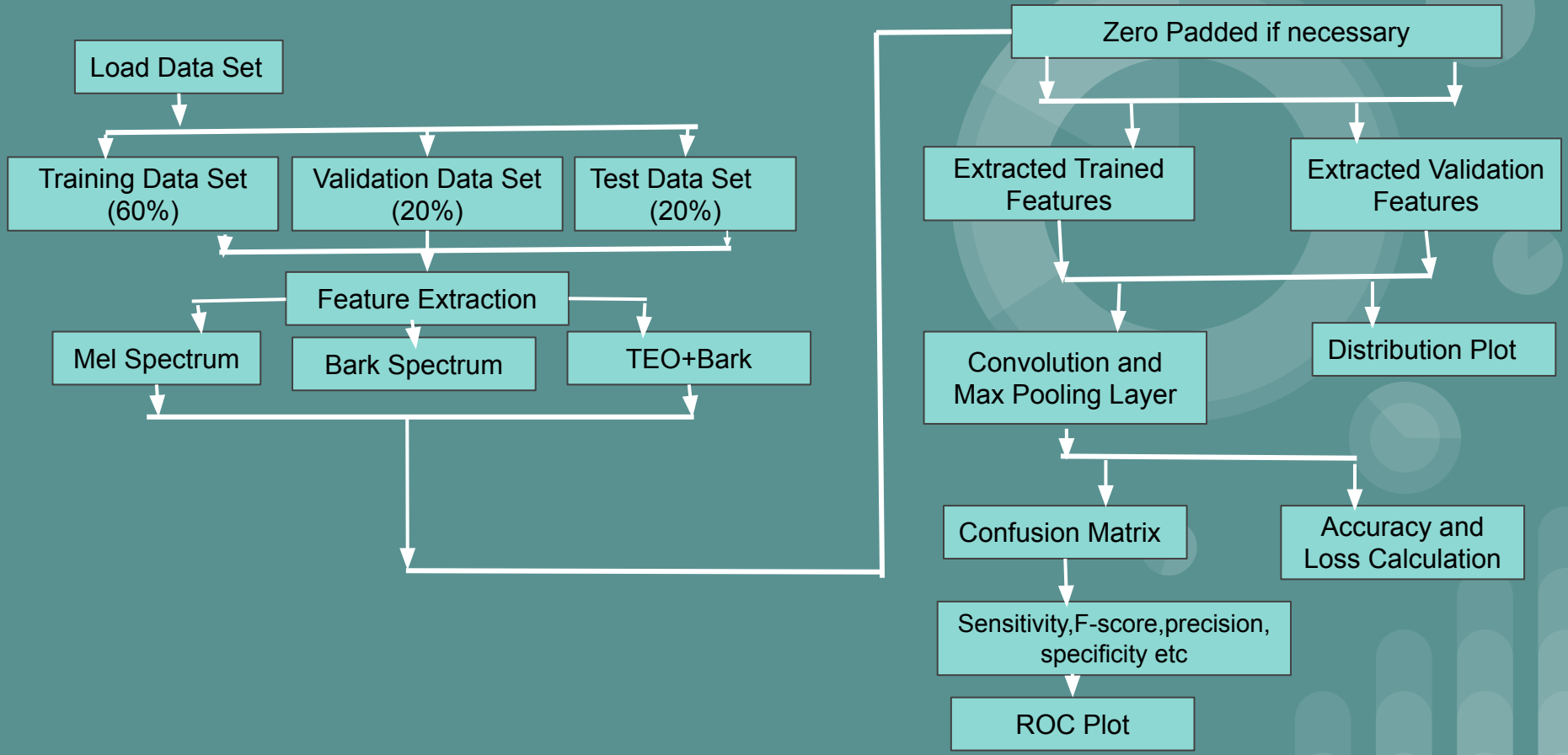
Augmentation were performed by shifting time and pitch,changing amplitude level,modifying tone,adding noise, stretching time etc, data was increased 10X times.

Tabular Summary of Male-Female Data:

Command Word	Male (53%)		Female (47%)		Total	
	Raw	Augmented	Raw	Augmented	Raw	Augmented
shamne	175	1750	154	1540	329	3290
pichone	185	1850	158	1580	343	3430
dane	168	1680	155	1550	323	3230
bame	172	1720	150	1500	322	3220
thamo	182	1820	148	1480	330	3300
Unknown	----------				6624	

Some recordings were affected by noise badly and they were not taken for training dataset,hence the number of raw data for each command is slightly different.

WorkFlow


Load Data Set
Training Data Set (60%)
Validation Data Set (20%)
Test Data Set (20%)
Feature Extraction
Mel Spectrum
Bark Spectrum
TEO+Bark
Zero Padded if necessary
Extracted Trained Features
Extracted Validation Features
Convolution and Max Pooling Layer
Distribution Plot
Confusion Matrix
Accuracy and Loss Calculation
Sensitivity,F-score,precision, specificity etc
ROC Plot

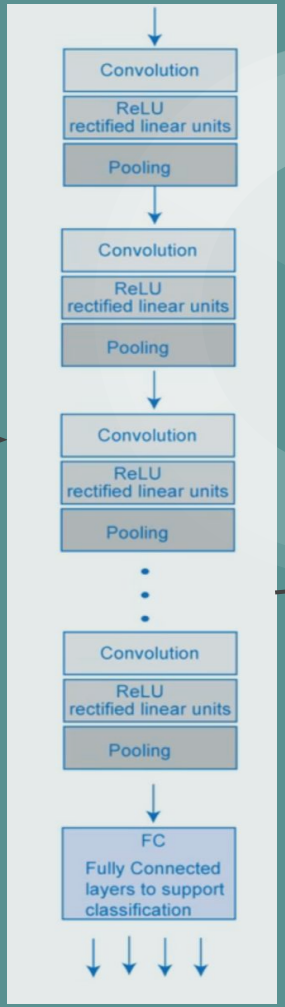
Model Architecture and Feature Information

Training and Validation dataset

Feature Extraction
Fs = 44.1kHz
2 sec Hanning Windowing
Frame Length = 0.025s
Overlapping = 0.015s

Num of features per frame = 50
Num of frames = 198

198x50 Input array per sample



5-Layer Convolutional Network with ReLU activation and max pooling (stride-3, padding-2) in between, followed by a fully connected layer

With loss function-**Weighted cross entropy**

6X1 prediction vector

Confusion Matrix using Test data set

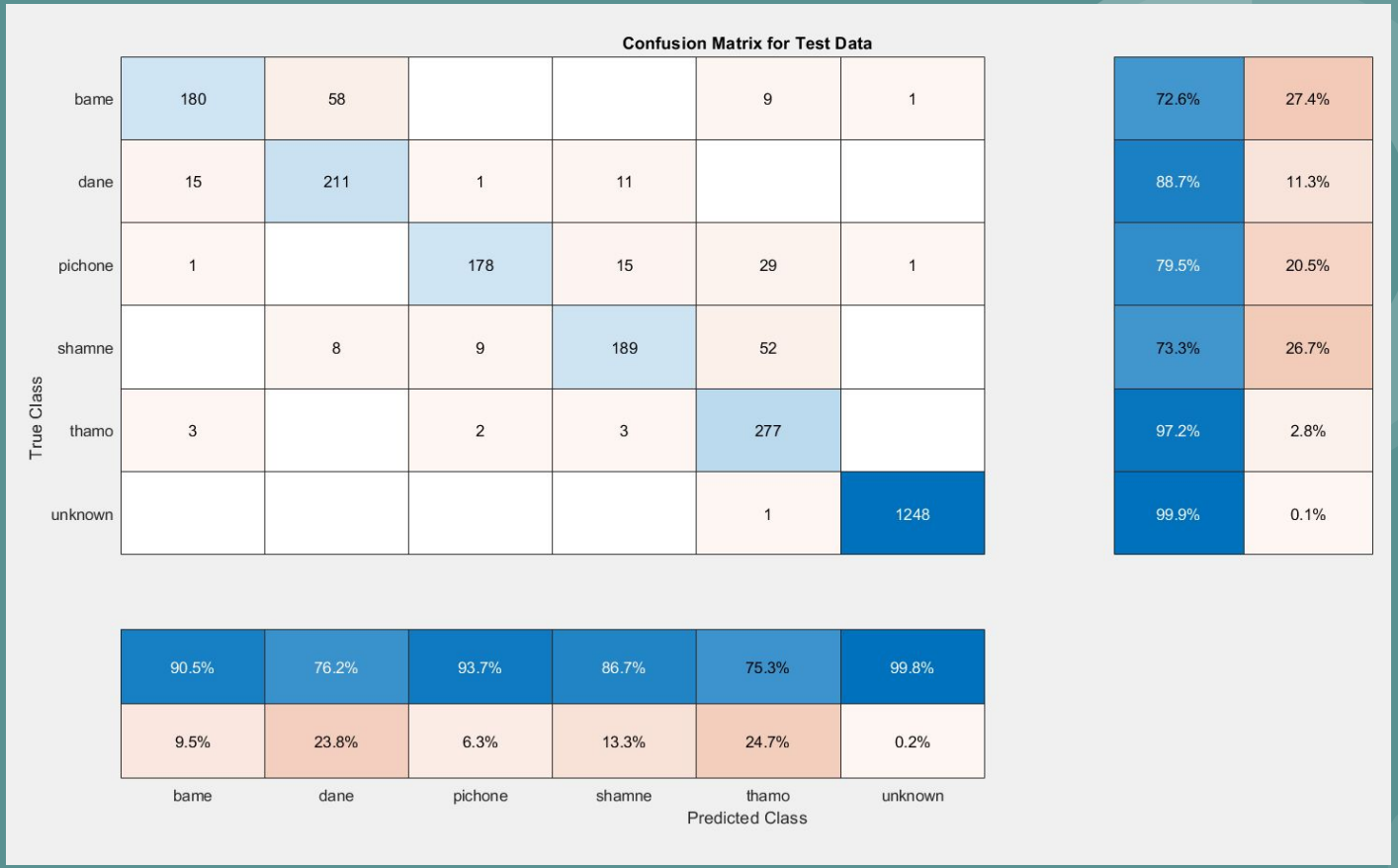
Choice of spectral descriptor

1. bark Spectrum
2. mel Spectrum
3. TEO+Bark

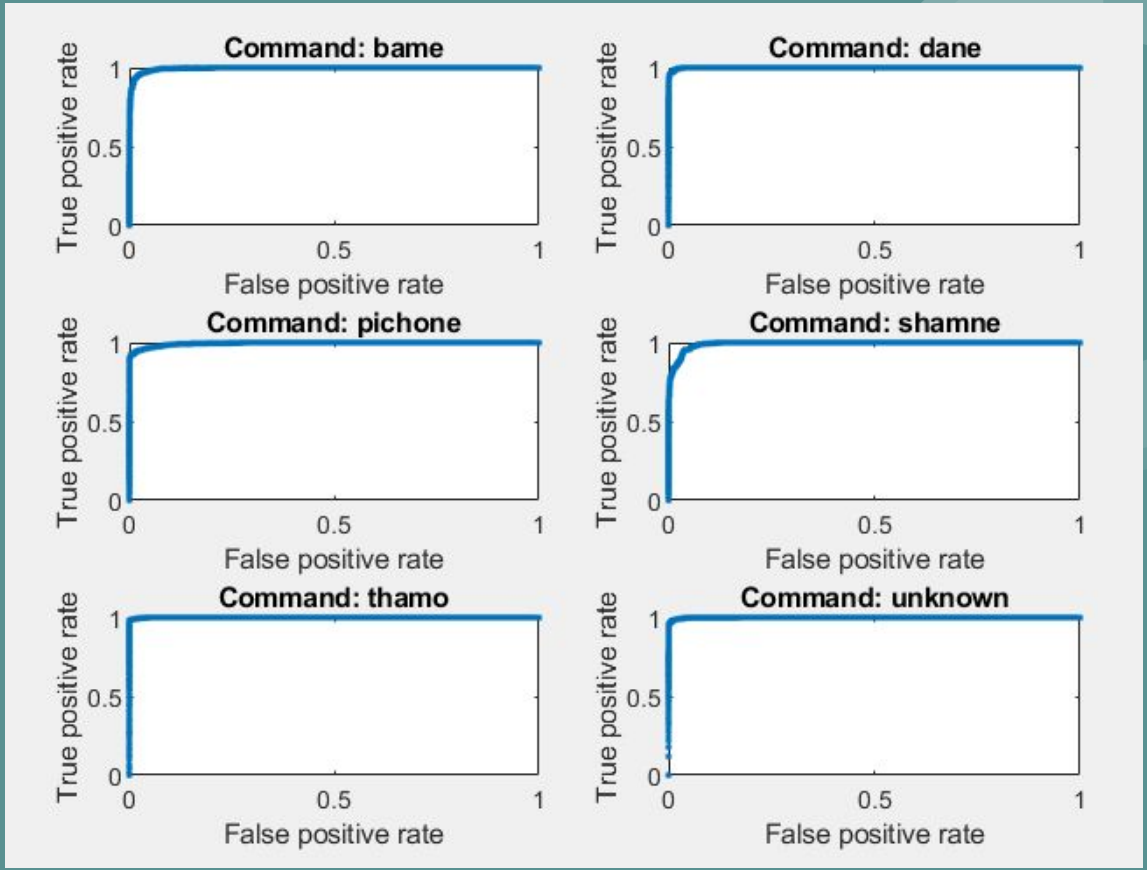
Result for Female Dataset Bark Spectrum:

Class Name	Bame	Dane	Pichone	Shamne	Thamo	Unknown
Accuracy	0.9652	0.9628	0.9768	0.9608	0.9604	0.9988
Sensitivity/Recall	0.7258	0.8866	0.7946	0.7326	0.9719	0.9992
Specificity	0.9916	0.9708	0.9947	0.9871	0.9590	0.9984
Precision	0.9045	0.7617	0.9368	0.8670	0.7527	0.9984
F-score	0.8054	0.8194	0.8599	0.7941	0.8484	0.9988

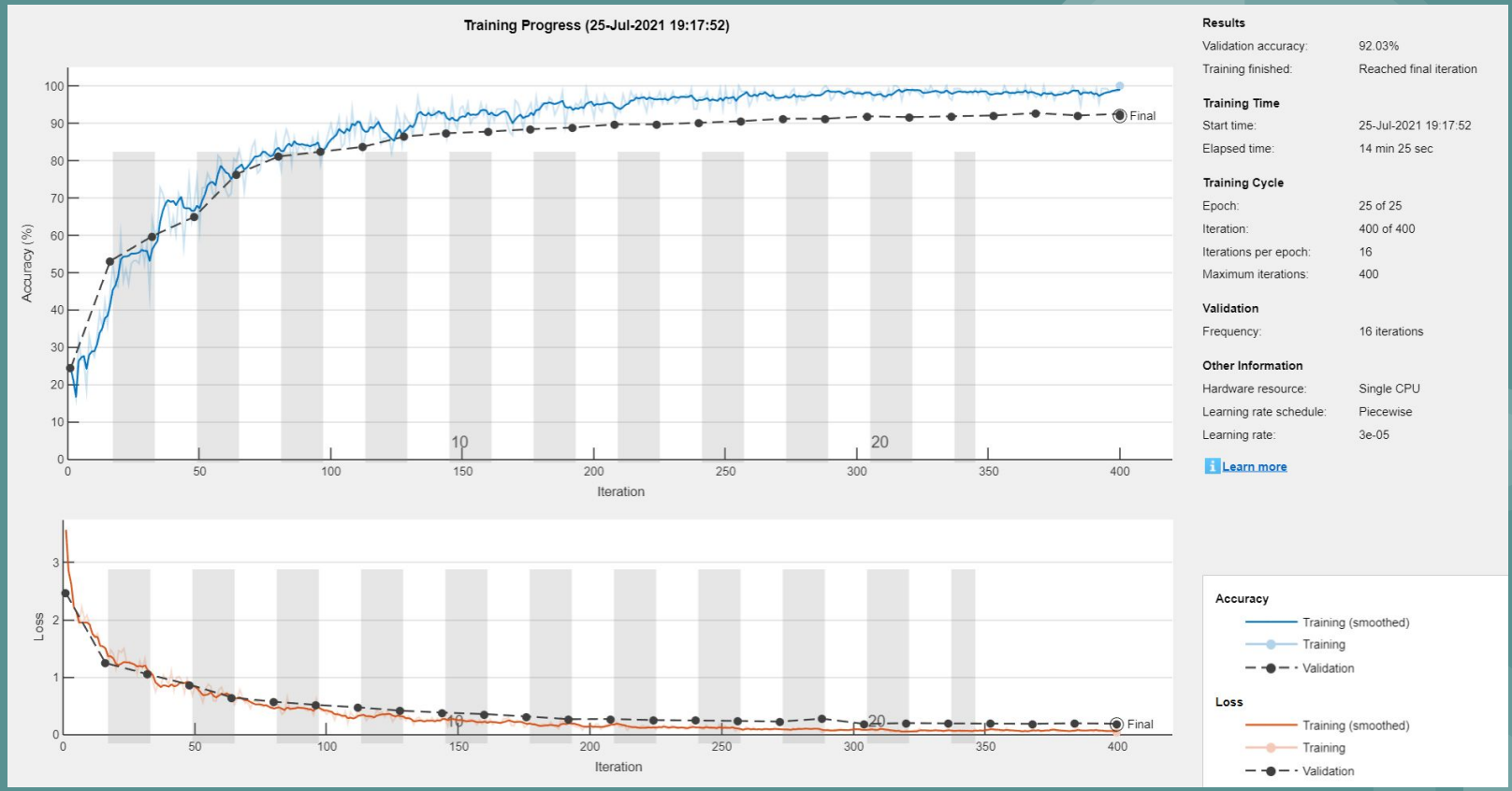
Result for Female Dataset Bark Spectrum:



Result for Female Dataset Bark Spectrum (ROC curve)



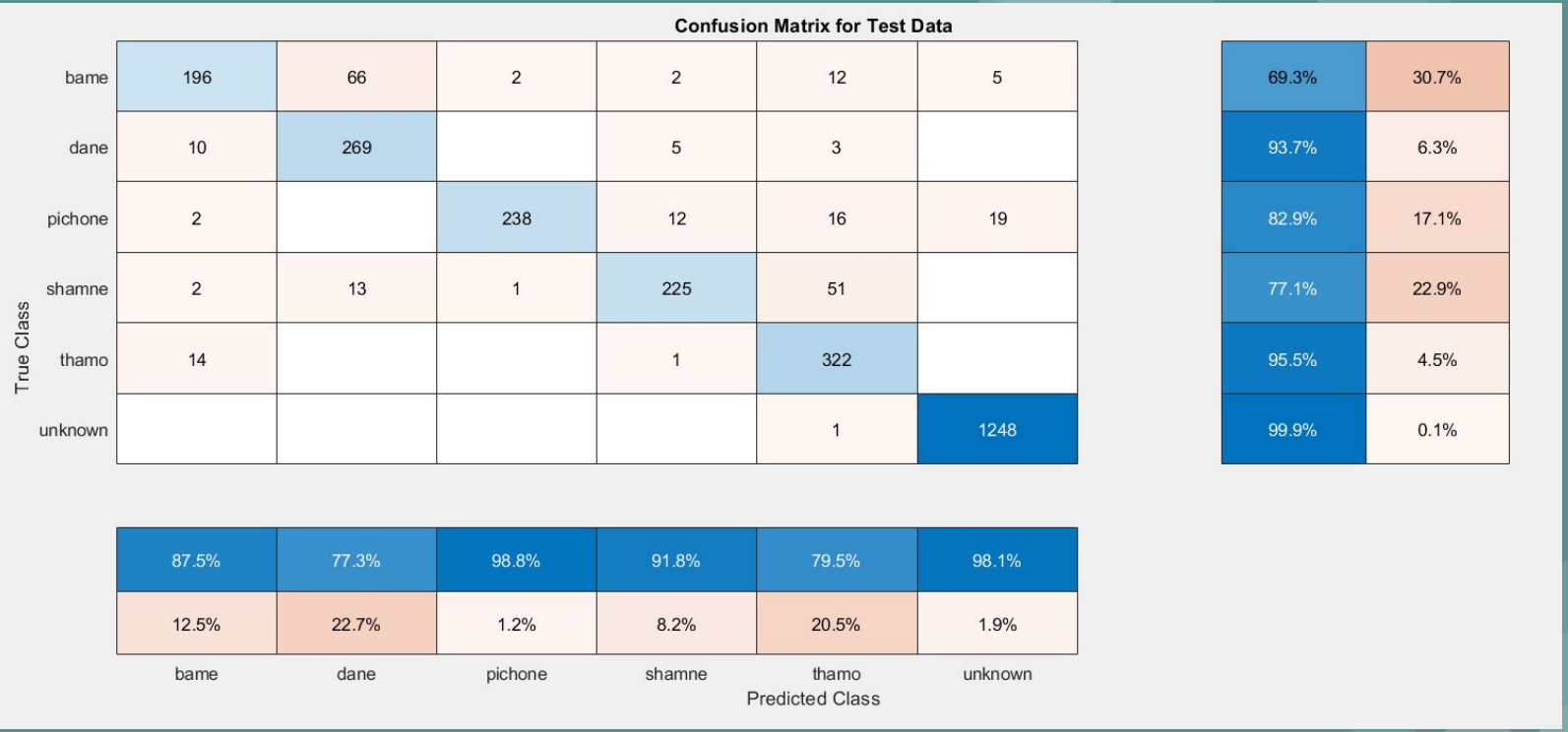
Result for Female Dataset Bark Spectrum:



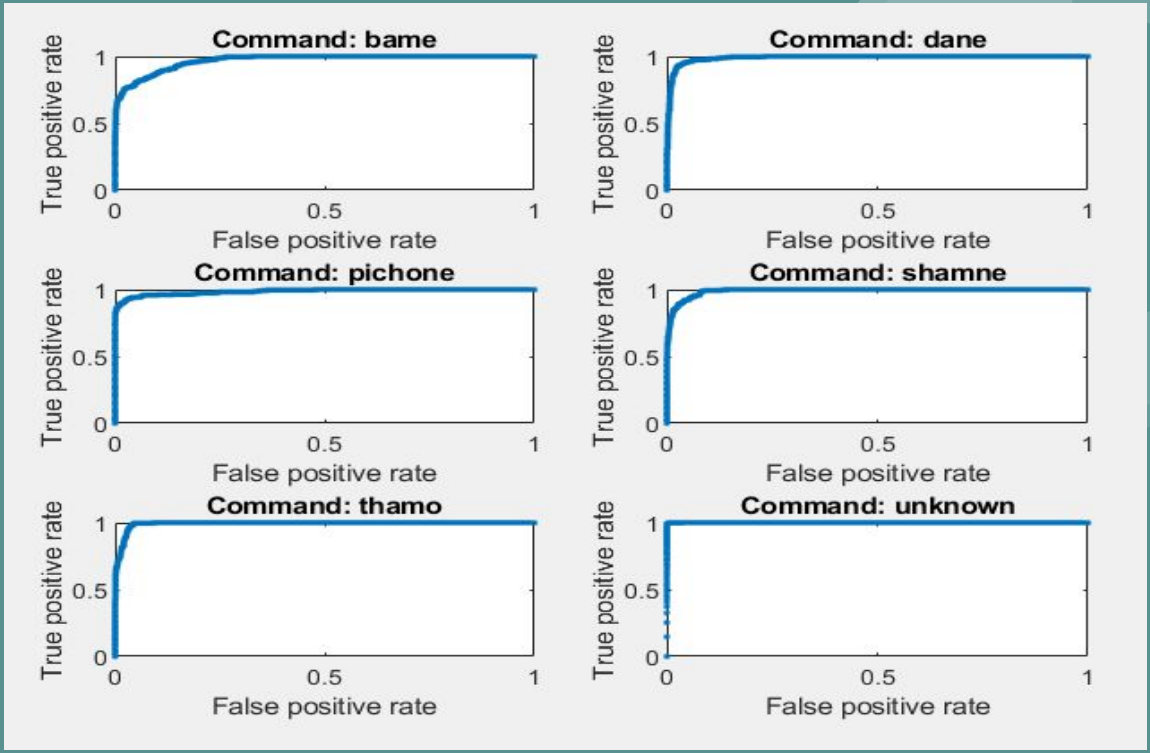
Result for Male Dataset Bark Spectrum:

Class Name	Bame	Dane	Pichone	Shamne	Thamo	Unknown
Accuracy	0.958	0.9645	0.9810	0.9682	0.9642	0.9909
Sensitivity/Recall	0.6926	0.9373	0.8293	0.7705	0.9555	0.9992
Specificity	0.9886	0.9677	0.9988	0.9918	0.9654	0.9838
Precision	0.8750	0.7730	0.9876	0.9184	0.7951	0.9811
F-score	0.7732	0.8472	0.9015	0.8380	0.8679	0.9901

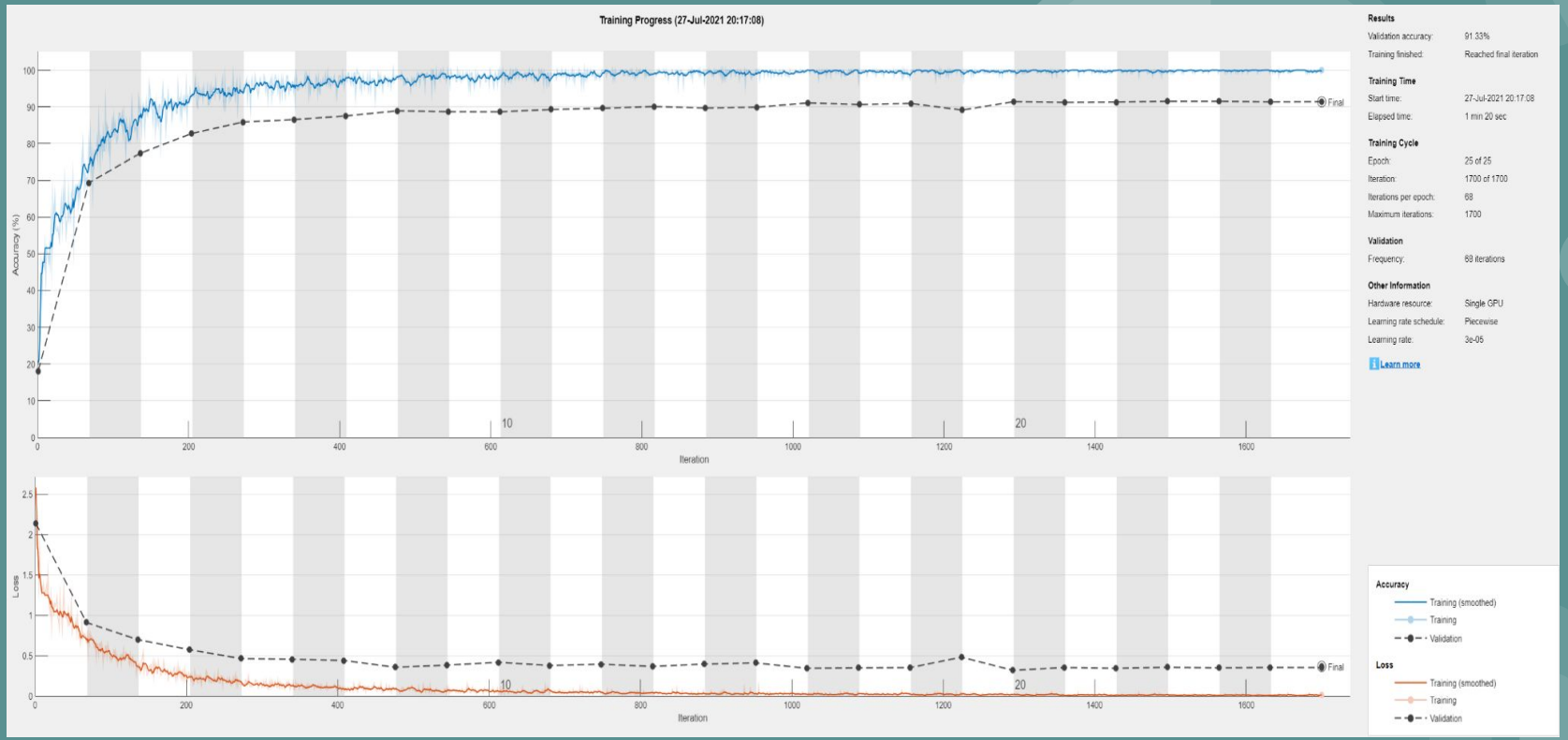
Result for Male Dataset Bark Spectrum:



Result for Male Dataset Bark Spectrum:



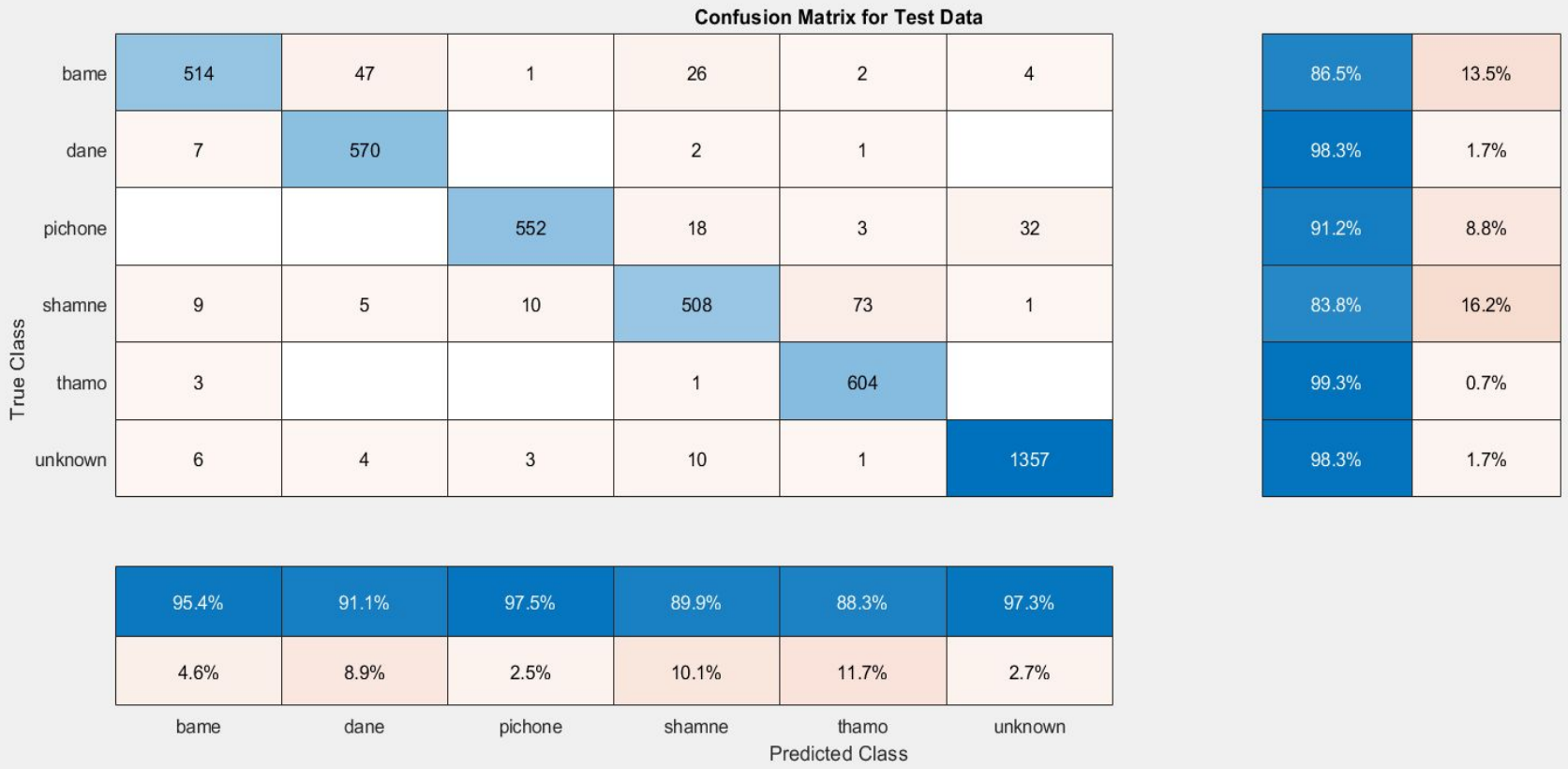
Result for Male Dataset Bark Spectrum:



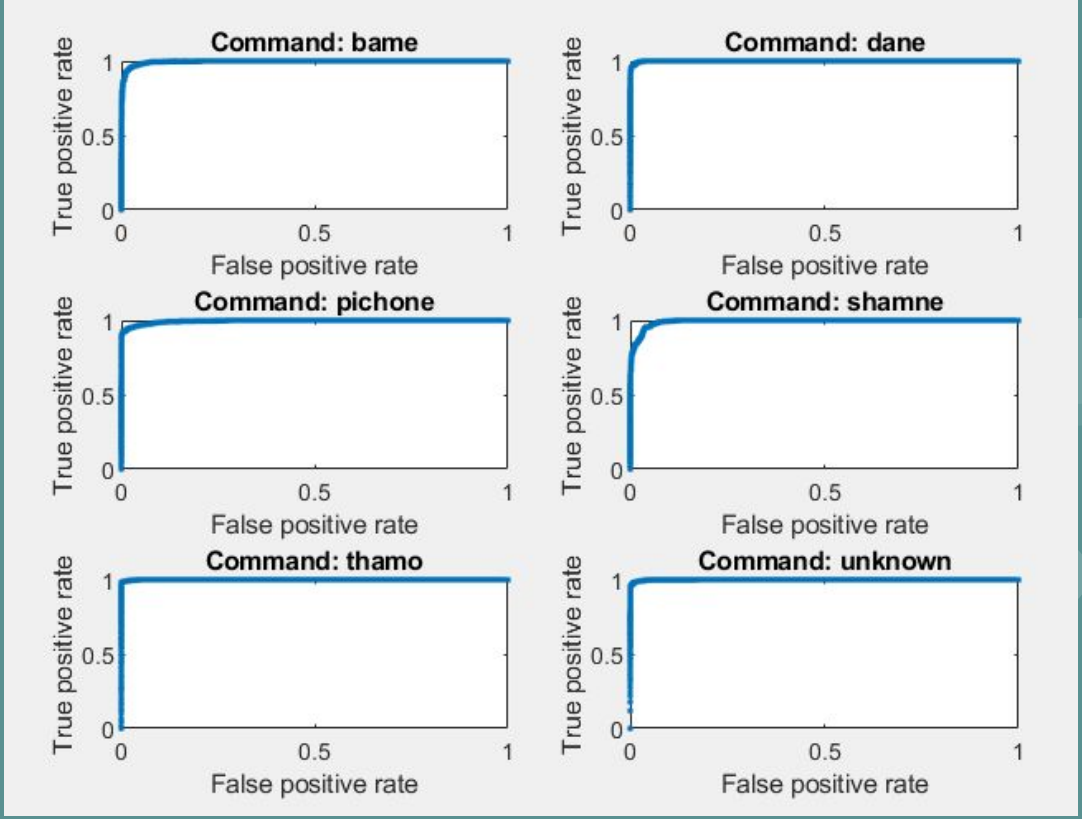
Result for Mixed Dataset Bark Spectrum:

Class Name	Bame	Dane	Pichone	Shamne	Thamo	Unknown
Accuracy	0.9760	0.9849	0.9847	0.9646	0.9808	0.9861
Sensitivity/Recall	0.8653	0.9828	0.9124	0.8383	0.9934	0.9826
Specificity	0.9934	0.9852	0.9963	0.9849	0.9788	0.9876
Precision	0.9536	0.9105	0.9753	0.8991	0.8830	0.9735
F-score	0.9073	0.9453	0.9428	0.8676	0.9350	0.9780

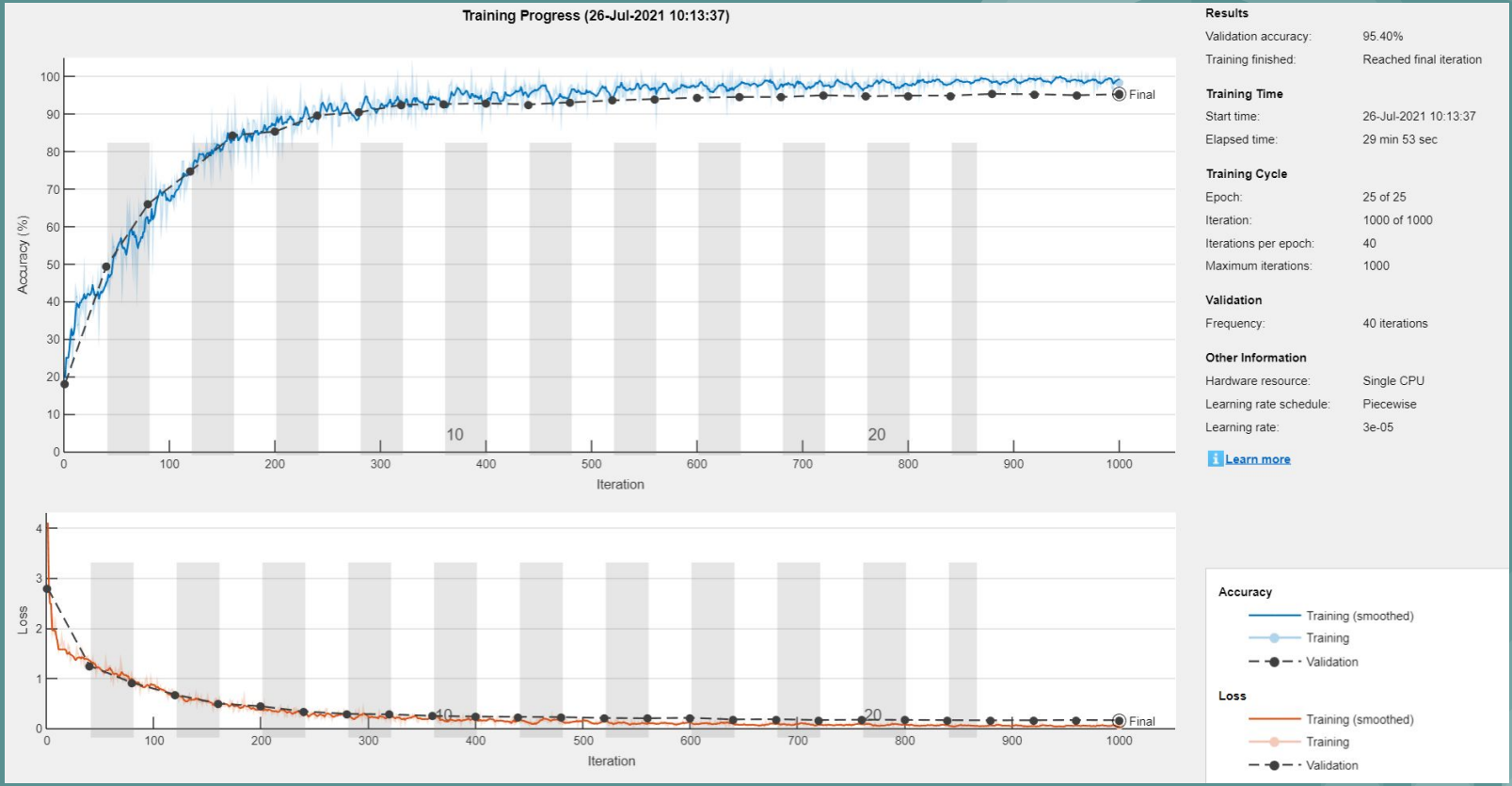
Result for Mixed Dataset Bark Spectrum:



Result for Mixed Dataset **Bark Spectrum**:

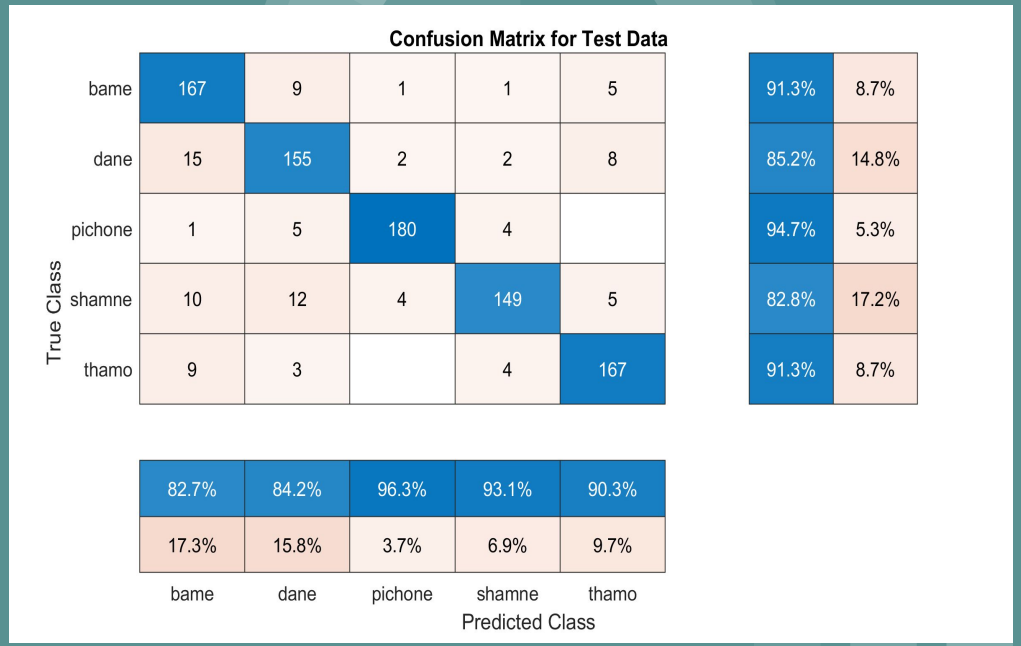


Result for Mixed Dataset Bark Spectrum:

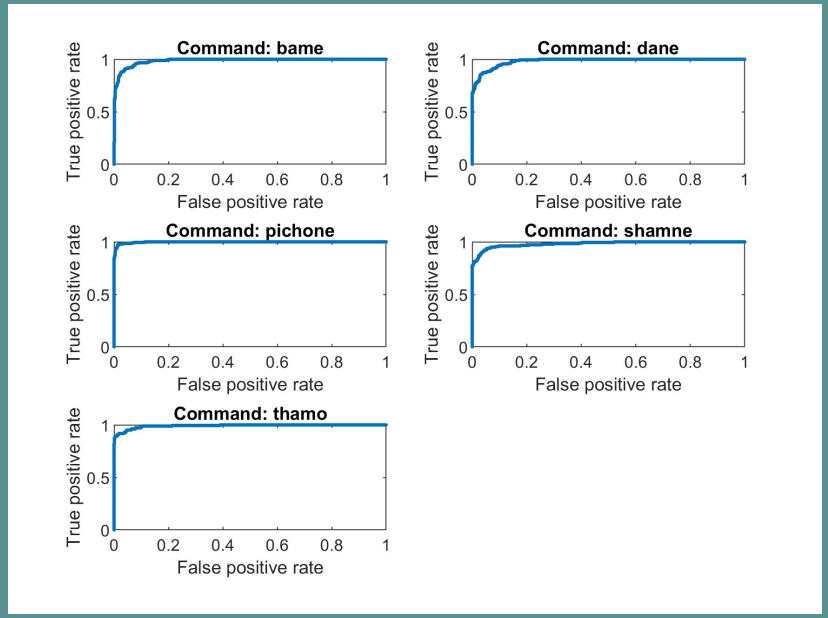


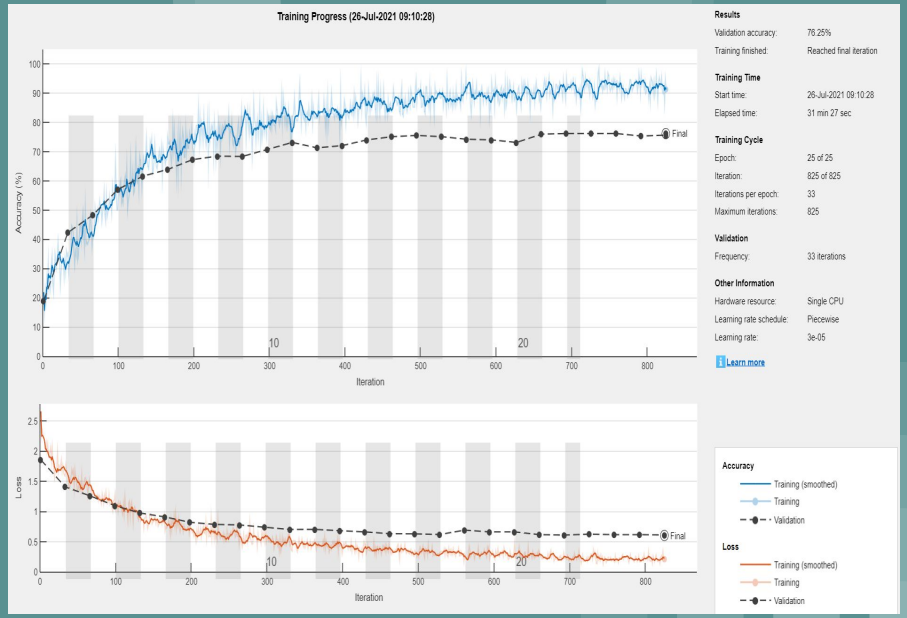
Result for Mixed Dataset With Teager Energy Operator + Bark Spectrum:

Class Name	Bame	Dane	Pichone	Shamne	Thamo
Accuracy	0.94	0.93	0.98	0.95	0.90
Sensitivity	0.91	0.85	0.94	0.82	0.91
Specificity	0.95	0.96	0.99	0.98	0.97
Precision	0.82	0.84	0.96	0.93	0.90
F-score	0.86	0.84	0.95	0.87	0.90

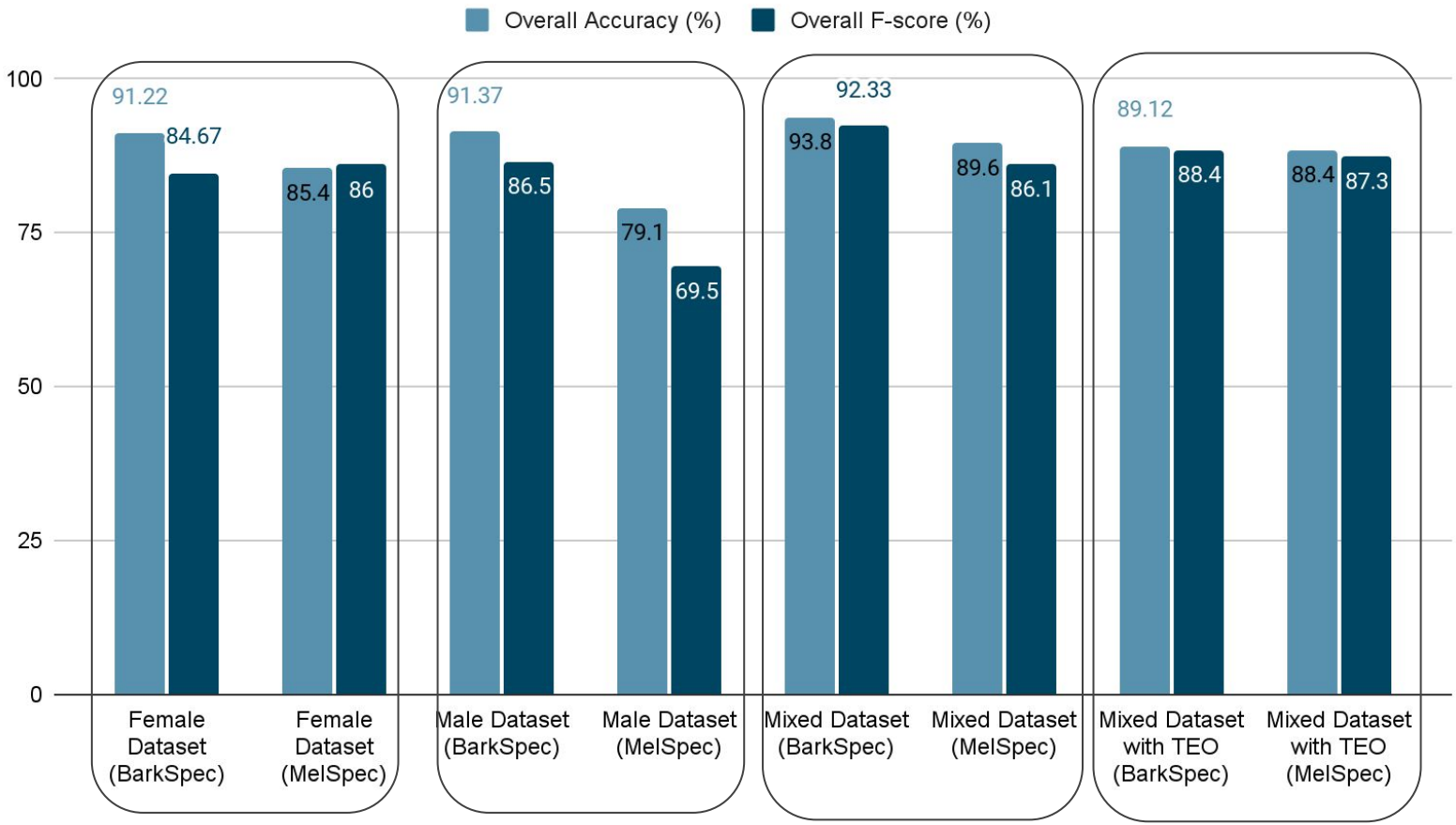


Result for Mixed Dataset With Teager Energy Operator and Bark Spectrum:





Comparison between different dataset for Mel, Bark and TEO+Bark spectrum



Reference Paper 1

Title:Bangla Short Speech Commands Recognition Using Convolutional Neural Networks

Output:

Model	Validation Accuracy(%)	Testing Accuracy(%)
MFCC	85.44	74.01
RAW	69.08	71.44
Transfer	68.06	73.00

Reference Paper : Bangla Short Speech Commands Recognition Using Convolutional Neural Networks, Shakil Ahmed Sumon, et. el, 2018 International Conference on Bangla Speech and Language Processing (ICBSLP)

Reference Paper 2

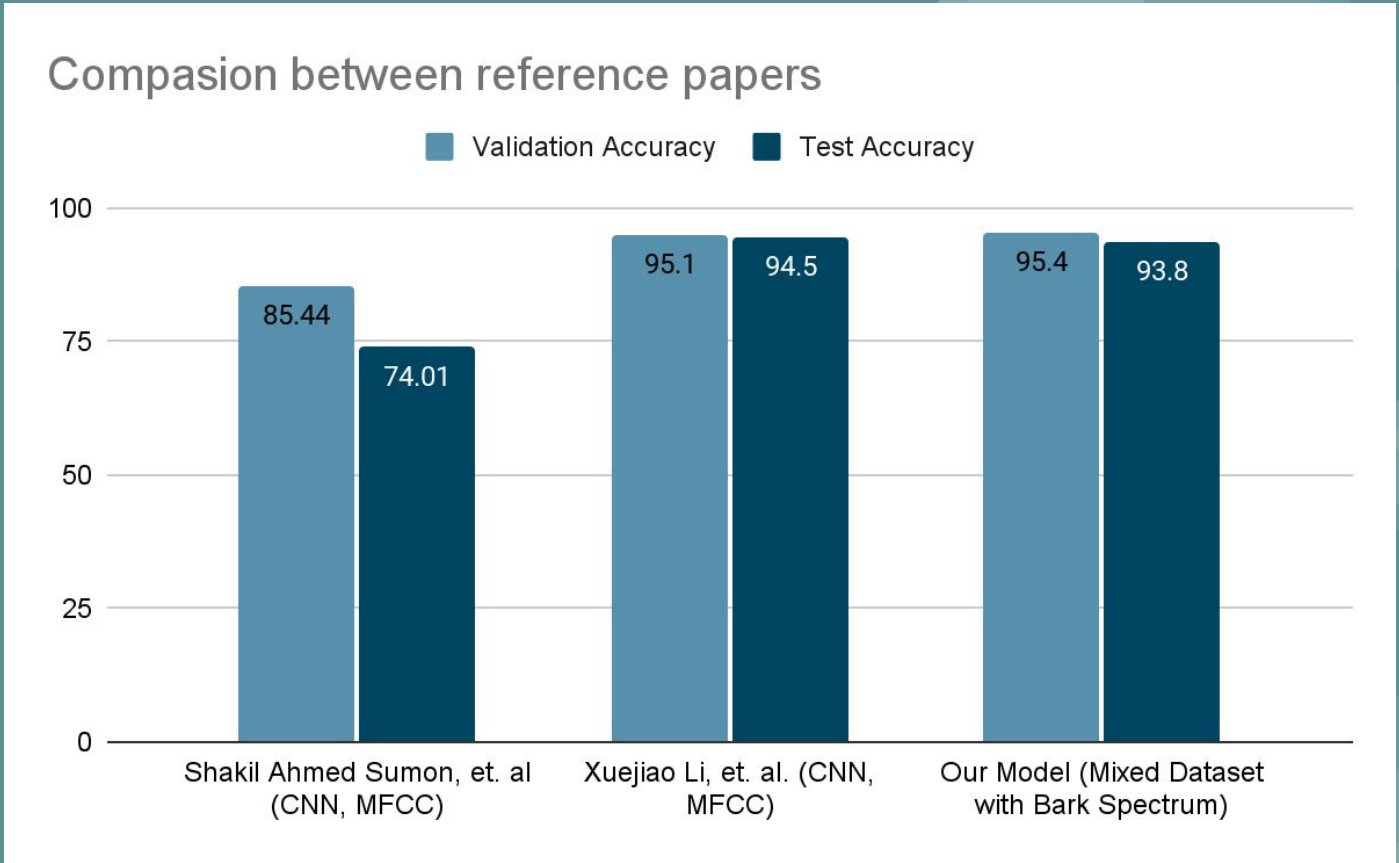
Paper Title:Speech Command Recognition with Convolutional Neural Network

Models	Validation Accuracy(%)	Test Accuracy(%)	Loss(%)
CNN	95.1	94.5	0.190
DNN	72.5	71.9	1.048
Vanilla	57.3	56.7	3.640

Keywords used: Go, Down, Up, Right, Left

Ref: Speech Command Recognition with Convolutional Neural Network, Xuejiao Li, et. el.

Comparison with reference paper

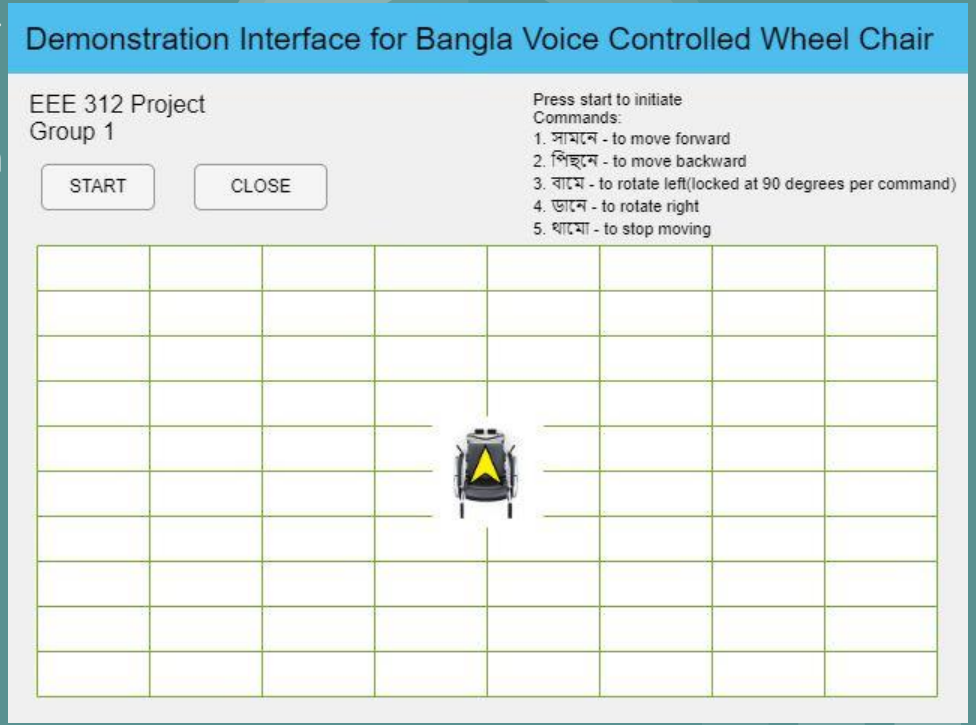


Model Implementation

- After training the model, we have designed an I/O system that takes voice input and gives corresponding command as output.
- Audio is taken in real time, and 2 second duration of audio is processed at a time for model input.
- For idle background, we used a power command that neglects input with very low power.
- Valid audio steam is updated and checked every 100 milliseconds.
- Since a command length is very less than 2 seconds, a command is detected multiple times throughout the 2 second window, and a memory system has been set up to check for outliers, and then rejected.
- Whenever a signal is detected multiple times (50% times of the memory), it is actually considered to be a command.

Virtual Demonstration Interface

- The interface shows a virtual wheelchair controlled by real time voice input from the user.
- The chair can perform linear motion with a constant velocity.
- The wheelchair can perform two kinds of rotation, continuous and discrete.
  - When asked to rotate, the chair rotates 90 degrees in given direction. If interrupted, the rotation can stop at any position.
  - To rotate at angles more than 90 degrees, another rotation command must be used.
  - In continuous mode, the wheelchair keeps on rotating until it is manually stopped by a command.



We look forward to calibrating our physical model for slopes and rugged grounds in future.

Demonstration Interface for Bangla Voice Controlled Wheel Chair

EEE 312 Project
Group 1

START CLOSE

Press start to initiate
Commands:
1. সামনে - to move forward
2. পিছনে - to move backward
3. বামে - to rotate left(locked at 90 degrees per command)
4. ডানে - to rotate right
5. থামো - to stop moving

Cost analysis

Sl. No.	Name of the Component	Quantity	Taka
1	Mechanical Wheelchair	1	4,000/=
2	Voice Recognition Module V3,1	1	9,000/=
3	Toyota Windshield Wiper Motor	2	2,000/=
4	Rotary Encoder	2	800/=
5	High Current Motor Drive	1	2,000/=
6	Scooter Motor	1	1,700/=
7	Raspberry pi	1	10,200/=
8	Microphone	1	900/=
		Total	**30.600/=**

Price Ref,- TechShopbd.com,store.roboticsbd.com

Future Possibilities

1. To add more commands to make this wheelchair more user-friendly.
2. To add sensor for bumpy and slope roads.
3. To differentiate between user and others to improve detection accuracy.
4. Optimize algorithm for better response time and less latency.

Thank you