

**Bangla-accented (Chittagong) Speech Recognition using Natural
Language Processing**

BY

**M Shahriar Ishtiaque
ID: 191-15-12938**

**Faraz Ahmed Bhuiyan
ID: 191-15-12948**

**Shazid Nawas Shovon
ID: 191-15-12929**

This Report Presented in Partial Fulfillment of the Requirements for the
Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Abu Kaisar Mohammad Masum

Lecturer

Department of CSE
Daffodil International University

Co-Supervised By

Mr. Md. Sadekur Rahman

Assistant Professor

Department of CSE
Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

FEBRUARY 2023

APPROVAL

This Project/internship titled “**Bangla-accented (Chittagong) Speech Recognition using Natural Language Processing**”, submitted by M Shahriar Ishtiaque, ID No: 191-15-12938, Faraz Ahmed, ID No: 191-15-12948 and Shazid Nawas Shovon, ID No: 191-15-12929 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfilment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 02/02/2023.

BOARD OF EXAMINERS

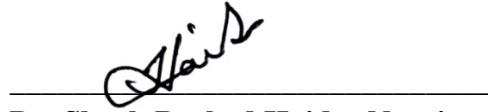
Chairman


6/2/23

Dr. Touhid Bhuiyan
Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

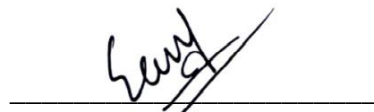
Internal Examiner



Dr. Sheak Rashed Haider Noori
Professor and Associate Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Md. Sazzadur Ahamed
Assistant Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner



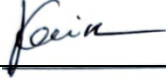
Dr. Md. Sazzadur Rahman
Associate Professor

Institute of Information Technology
Jahangirnagar University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Abu Kaisar Mohammad Masum**, Lecturer, **Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



Abu Kaisar Mohammad Masum

Lecturer

Department of CSE

Daffodil International University

Co-Supervised by:



Mr. Md. Sadekur Rahman

Assistant Professor

Department of CSE

Daffodil International University

Submitted by:




M Shahriar Ishtiaque

ID: 191-15-12938

Department of CSE

Daffodil International University



Faraz Ahmed Buiyan

ID: 191-15-12948

Department of CSE

Daffodil International University



Shazid Nawaz Shovon

ID: 191-15-12929

Department of CSE

Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We are really grateful and wish our profound our indebtedness to **Abu Kaisar Mohammad Masum, Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of machine learning, natural language processing and data mining to carry out this project to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Professor Dr. Touhid Bhuiyan, Head, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Speech recognition has been a very popular topic for the students of computer science and engineering. Over time, with the advancement of technology, we can now recognize Bengali speech as well. But people from all corners of Bangladesh do not speak the standard language. Therefore, sometimes it becomes difficult to recognize the local accented speech. With the hope of resolving this issue, we have made every effort to develop a local-accented Bengali speech recognition system. In our system, we have detected the audio with a CNN (convolutional neural network) by converting audios into a spectrogram. Aside from audio recognition, it also provides sample data that is simple to handle and use. Then we have done gradient boosting with MFCC (Mel Frequency Cepstral Coefficients) features of audio data, which reduces the prediction error by combining with the previous model. It is important to note that MFCC is a popular algorithm used to filter vocal tracks. Using this, we classified the local accented speech sample, which is further assured by the Random Forest Classifier. Finally, we got the desired output from the ANN (artificial neural network) model. For checking the functioning of our system, we have been using the local accent of the people of Chittagong as a reference. We got maximum 91% training accuracy and 81% test accuracy with 3161 audio files of 25 classes. We are hopeful to expand our research on this topic, which will be very helpful to the people who speak in a local tone to be at par with others

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
CHAPTER	
CHAPTER 1: INTRODUCTION	1-6
1.1 Introduction	1
1.2 Motivation	3
1.3 Rationale of the Study	4
1.4 Research Question	5
1.5 Expected Out come	5
1.6 Report Layout	5
CHAPTER 2: BACKGROUND	7-10
2.1 Related Works	7
2.2 Comparative Analysis and Summary	8
2.3 Scope of the Problem	9
2.4 Challenges	9
CHAPTER 3: RESEARCH METHODOLOGY	11-20
3.1 Research subject and instrument	11
3.2 Dataset utilized	12

3.3 Statistical Analysis	16
3.4 Methodology	17
3.5 Implementation Requirements	18
Chapter 4: Experimental Results and Discussion	21-30
4.1: Experimental setup	21
4.2: Experimental Result and analysis	29
Chapter 5: Impact on Society, Environment and Sustainability	31-32
5.1: Impact on Society	31
5.2: Impact on Environment	31
5.3: Ethical Aspects	31
5.4: Sustainability Plan	32
Chapter 6: Summary, Conclusion, Recommendation and Implication for Future Research	33-34
6.1 Summary of the Study	33
6.2 Conclusion	33
6.3 Implication for Further Study	34
REFERENCES	35-36

LIST OF FIGURES

FIGURES	PAGE NO
Figure 2.1: In General Process	8
Figure 3.1: Flowchart of audio to spectrogram conversion algorithm	12
Figure 3.2: Steps of generating MFCC	13
Figure 3.3: Wave Plot	13
Figure 3.4: Generation of Mel Spectrogram	17
Figure 3.5: Total Data of Each classes	17
Figure 3.6: Methodology Figure	16
Figure 4.1: Normalized confusion matrix of CNN	22
Figure 4.2: train accuracy vs validation accuracy curve	22
Figure 4.3: train loss and validation loss	22
Figure 4.4: Actual and predicted classes and confidence percentage of CNN model	23
Figure 4.5: Confusion matrix of Gradient boosting classifier model	24
Figure 4.6: Train accuracy vs Validation accuracy curve	25
Figure 4.7: Train loss and Validation loss	25
Figure 4.8: Normalized Confusion matrix of Random Forest classifier	25
Figure 4.9: Train accuracy vs Validation accuracy curve	26
Figure 4.10: Train loss and Validation loss	26
Figure 4.7: Confusion Matrix of ANN	27
Figure 4.8: train accuracy vs validation accuracy curve	27
Figure 4.9: train loss and validation loss	27
Figure 4.10: ANN model summary	28

LIST OF TABLES

TABLES	PAGE NO
Table 3.1: INFORMATION OF DATASET	16
Table 4.1: Comparison table of 4 classifiers	29

CHAPTER 1

Introduction

1.1 Introduction:

Research has been conducted for decades with the goal of recognizing and synthesizing accents; yet the automated speech recognition system (ASR) was not established until the middle of the twentieth century. The Automatic Speech Recognition (ASR) system is utilized in natural language processing (NLP) for the purpose of recognizing speech. NLP is a subfield of artificial intelligence (AI) that utilizes computational methods and linguistics in order to analyze natural languages, such as those spoken by people. This subfield is extremely important since it plays a critical part in the field. The scope of what is meant by the word natural language processing (NLP) has been broadened to encompass a number of subfields within the fields of science and technology, such as the classification of text, classification of pictures, voice identification, etc. The capacity of a computer software to convert spoken words into written language is referred to by a few different names, including automatic speech recognition (ASR), computer voice recognition, and speech-to-text. A speaker has to "train" (also known as "enroll") a voice recognition system by reading text or a restricted vocabulary to the device. This process is also known as "enrolling" the system. Another term for training is "enrollment," and both terms are used interchangeably. The individual's voice is evaluated by the system, and the results of that analysis are combined with other data in order to improve the accuracy of the speech recognition. The functioning of systems that are referred to be "speaker-independent" does not require any training to be performed. Systems that are based on training are referred to be "speaker dependent" when using the word "speaker dependent" to describe them [10].

Speech recognition has been around for a while and has seen a lot of technical progress over the years. The subject of artificial intelligence has advanced thanks to recent advancements in deep learning and large data. There has been an uptick in the number of scholarly articles on the subject, but perhaps more significantly, companies across

the world have begun using a variety of deep learning approaches in the development and deployment of their own speech recognition systems. One of the most debated subjects in the field of NLP is the identification of accent. The accent of a language is a representation of the people who live in a certain place as well as the social or economic class to which they belong. Several approaches to automatic accent recognition were examined, each of which used a diverse set of ideas taken from various other facets of speech technology. Previous work has been done in an effort to build automatic voice identification systems. This is due to the fact that considerable levels of accent diversity within each language can be challenging for a system to deal with [11].

We anticipate that variation associated with accents will be distinguishable from speaker-to-speaker variation, variation associated with disfluency, and other types of variation associated with speech distortion. When compared to speech artifacts, an accent is more likely to maintain its original form over the entirety of a sentence. The vast majority of dialects, on the other hand, may be classified linguistically (in accordance with the speakers' first languages), which is in stark contrast to the wide range of speakers. The control of accent variation is a greater challenge than the following categories of variation: Even native speakers need to be exposed to dialects that are quite close to their own in order to be able to comprehend speech delivered with a distinctive accent. Because of this, it makes perfect sense to teach a neural network to recognize accented speech by acclimating it to a wide range of different speech inflections. Bengali is the local language that is spoken by the fifth most people, and it is the seventh most spoken language in the entire globe. There are about 228 million people in the globe who speak Bengali as their first language, and an additional 37 million people use Bangla as a second language [14]. There are a lot of local Bengali dialects that are easily distinguishable from one another. People in all parts of Bangladesh speak one of eight primary regional dialects: Dhaka (old), Khulna, Chittagong, Mymensingh, Sylhet, Barisal, Rangpur, or Rajshahi. These are the names of the cities where these dialects are spoken.

Spoken in the Chittagong Division of Bangladesh, Chittagong (sagia or siaiga) is an Indo-Aryan language. Chittagong is spoken by people who consider themselves

Bengali, however the language is not mutually understandable with Bengali. Many linguistic experts classify Chittagong as a distinct tongue. It shares basic comprehension with Rohingya and, to a lesser extent, Noakhailla. Chittagong is spoken by between 13 and 16 million people, mostly in Bangladesh, according to a 2009 estimate.

The Indo-European language family includes Chittagong, which belongs to the Bengali-Assamese sub-branch of the Eastern group of Indo-Aryan languages. It is a descendant of Proto-Indo-European and Old Indo-Aryan through an Eastern Middle Indo-Aryan. Grierson (1903) classified Chittagong's dialects with those of Noakhali and Akyab as Southeastern Bengali. According to Chatterji (1926), Chittagong belongs to the Vangiya group of Magadhi Prakrit, and all Bengali dialects developed separately from the literary Bengali known as sadhu bhasha. Chittagong, one of the several varieties of eastern Bengali, contains phonetic and morphological features that are distinct from standard Bengali and other western varieties of Bengali [15].

This study produces an application that can estimate a person's geographical location based on the frequency of their voice waves by utilizing both the standard Bangla language and the regional Bengali dialect that is utilized in the Chittagong area. The research was conducted in Bangladesh. In this particular instance, MFCC was utilized for the purpose of feature extraction, and a number of techniques were utilized in order to attain the highest possible level of system accuracy.

1.2 Motivation:

There is a plethora of modern conveniences accessible only by the power of the human voice. One of the most vital areas of NLP is speech recognition. Voice recognition has been the subject of several works. Although artificial intelligence (AI) assistants like Siri [17], Alexa [18], etc. have been developed, they are not yet capable of accurately identifying regional or linguistic variations in speech. Unfortunately, there is a dearth of resources for the Bengali language, and the available systems are only of low quality. As Bangla is one of the most widely spoken languages in the world and has a rich cultural heritage, a reliable speech recognition system for Bangla is essential. There are several regional dialects spoken in Bangladesh. Most Bangla-supporting AI assistants

are designed just for standard Bengali. Those living in the countryside who don't speak standard Bangla well would have trouble accessing the services provided by these networks.

Because of this, we set out to develop a system that could identify different dialects spoken in Bengal based on their spectral characteristics. The study's overarching goal is to develop, using a robust dataset, a system that facilitates human-computer interaction for the benefit of those living in rural areas, with the highest possible degree of precision.

1.3 Rationale of the Study

The processing of natural languages has been the focus of a significant amount of research; nonetheless, the great majority of these studies have been carried out and published in English. These methods or processes are implemented in a number of different automated frameworks. Although there are around 300 million Bangla speakers across the globe, only a tiny number of academics are actively pursuing the study of this language, and the majority of those who do have a poor level of accuracy in their language skills. In addition, the spoken dialects of Bangladesh include a variety of regional differences that may be found throughout the country. Speech patterns may be used as a reliable predictor of a person's place of origin thanks to the myriad of ways in which they might vary. The domain of natural language processing (NLP), specifically the subfield of Bengali accent recognition, suffers from a serious lack of advancement.

In the present day, voice-based apps have become increasingly important. These applications have the potential to aid society in a variety of ways, including as an AI assistant, in census surveys, in the fields of finance, human resources, marketing, medicine, transportation, auto-typing, detecting crimes, and more. Our interest in this field stems from a desire to see voice-based apps in Bengali dialect used more extensively for the benefit of rural populations.

1.4 Research Questions

How can we collect the data of Chittagong dialects?

- What techniques can be used for feature extractions?
- Which classification algorithm should be used to get the best outcome?
- Can Bangladesh's rural residents benefit from the findings of this study?
- Do vocal folds that represent the same language with Chittagong accents produce different wavelengths?

1.5 Expected Outcome

This study's eventual outcome aims to create a tool that will let the average citizen of Bangladesh learn about the Chittagong people and their old language and begin to comprehend what they are saying. There may be a lot of uses for an app like this. Things like:

- Numerous types of offenses are committed through phone calls and audio recordings, and they can be identified by their accents and proven in court.
- The tourist attractions and international automatic telephone services could benefit from this study because they could use it to determine a person's region based on their accent.

1.6 Report Layout

In Chapter 1, This paper addresses not only what we want to accomplish, but also why we intend to do it, and how we intend to achieve it. This chapter provides a concise explanation of the goals that guided the creation of this work as well as the outcomes that were anticipated.

In Chapter 2, Tasks associated with this industry's job have been outlined. In addition, a summary of the research results is included. We determine what to aim for by first learning what we can't do.

In Chapter 3, The paper includes a discussion of the research methods that were employed over the course of this study. In this chapter, we take a superficial look at the data gathering procedure, data preparation, feature extraction, and approaches employed in this study.

In **Chapter 4**, This chapter presents the outcomes of the preceding chapters as well as comparisons and the best processes.

In **Chapter 5**, The project summary is the major emphasis. This report's last chapter includes a discussion of future work, a conclusion, limitations, and recommendations.

CHAPTER 2

Background

This section will discuss previous works that have inspired our own. Little academic work has been done on the topic of identifying and categorizing words with a particular rural accent. Different languages, of course, have unique pronunciations. Many studies have been conducted on the basis of these accents. Our study, however, contrasts with the old categorization studies. In this chapter, we shall summarize the findings of those studies. The research's strengths and weaknesses, as well as the project's overall framework, will also be reviewed.

2.1 Related Works

Before beginning our thesis, we conducted extensive study and accumulated a large amount of data. We've read a lot of paper and articles on our issue.

Recognizing accent from language is not another examination any longer. Numerous analysts are as of now attempted to recognize emphasize in numerous ways. In this paper we utilized many Machine learning and Deep learning Algorithm to sort out the most ideal exactness.

In study [1], the researchers employed a local naive Bayesian nearest neighbor method, which involves analyzing spectrogram images of speech signals, to categorize isolated speech sounds using Scale-invariant Feature Transform (SIFT) features (LNBNN). This technique could potentially be used to classify a group of features with diverse shapes and sizes.

In several research papers [2],[3],[4] image processing techniques are used to create a spectral image of speech sounds and then utilize CNNs (convolutional neural networks) to determine the phonemes being spoken. In another paper [5], the authors discuss methods for recognizing emotions using image processing techniques. Additionally, the use of image processing to identify speech signals through spectral imaging is described [6].

In [4], the authors propose an adaptive recognition system for different accents in conversations using a CNN (convolutional neural network). This system is intended to

be used in a call center setting to address dialogue speech recognition issues involving different accents. Other studies have used deep CNNs to analyze patients' EEG signals and determine the degree of similarity with clinically diagnosed symptoms [7]. The author of [8] says that the environmental sounds are converted into spectrogram images and classified using a CNN. In [9] presents a new method for extracting features for sound event classification, which is based on a visual signature extracted from the time-frequency representation of the sound.

So, this method is notable because spectrograms generally contain more information than the hand-crafted features that are commonly used for audio analysis.

2.2 Comparative Analysis and Summary

A lot of research has been done on accented language classification, some of them used MFCC, LSTM or other feature extractor or some research used image approach by using spectrograms. Some research work used noise free environment for data collection but some of them did it normally on a noisy environment but ultimately, they shaped the data on a different shape but before that people who have noisy data used coding approach to remove the noise. Even though the data quality of both of the research were not the same but preprocessing of them gave almost same result. Because most of the data was used for training purpose and less data was used for test purpose.

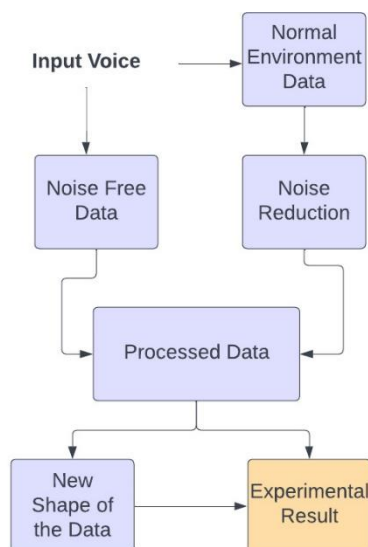


Figure 2.1: In General Process

2.3 Scope of the Problem

In our study Chittagong Accent detection was done. Like Chittagong Accent detection any kind of Accent detection may help in a lot of ways. Identification of perpetrators or the location of the crime's most dangerous area may be useful in some circumstances. For example, using this accent categorization technique, we may identify the region from where a criminal originates if we heard a voice. And in other circumstances, this may be the only hint that a crime has a chance of being successful. Also, Chittagong district is a tourism spot of our country. So, our model implementation in a system can help the tourist them to communicate with the locals effortlessly.

However, this system can also aid in marketing. The region of a businessman's targeted consumer can be determined with this approach even if they only have a small number of customers in mind. He or she can turn a profit through this. So, there's no doubt that this approach aids businessmen in locating their target market.

2.4 Challenges

1. Data Collection

One of the primary difficulties we faced was collecting the necessary data for our project. Specifically, we found it challenging to obtain organized datasets for the Bangla language. Furthermore, it was difficult to find a sufficient number of sources for data on Chittagong accents of the language. In order to gather the necessary data, we had to travel to Chittagong as it was not possible to collect data on the authentic Chittagong accent in a Dhaka City. Another challenge we encountered was the need for specialized data, such as music-free and noise-free voices, which proved difficult to obtain. To overcome this issue, we had to go to Chittagong and collect some data and plant a representative there for the further gathering of more needed data.

2. Collect the exact Accent

Previously, we mentioned that one of the main challenges for this project was obtaining the necessary data. Another issue we faced was that people often mix different accents when speaking, which made it necessary for us to reject certain voices or speech samples. This made it more difficult for us to find suitable data for our project. So, for the authentic Chittagong accent We had go to Chittagong and collect the exact accented voice.

3. Ensuring that language is compatible with the system

Parsing and structuring Bangla text was a significant challenge due to the complexities of the grammar, which is more intricate than that of English. This made it difficult to properly format the data and ensure it was compatible with the system. However, we persevered and through our efforts, we were able to gain a deeper understanding of the language. Despite the difficulties we encountered, the process of researching and learning about Bangla has been a rewarding experience.

4. Model Selection

Despite the significant amount of research that has been conducted on accent classification, it remains a challenging area of study for the Bangla language. Much of the research on this topic has been conducted using the English language, resulting in the development of numerous models for accent classification. However, it has been difficult to identify the most appropriate model for the Bangla language that will provide the highest level of accuracy.

5. Workflow

In order to determine the most efficient and effective workflow for this project, we conducted various experiments and tested out different processes. After analyzing the results of these experiments, we were able to identify the workflow that produced the best results. This workflow is outlined below:

Chapter 3

Research Methodology

3.1 Research subject and instrument

In this section, we will discuss the methods and instruments used to collect data. Our research consists of using Neural network algorithm. That's why we needed good amount of

Audio data to train the neural network model. The research is about speech detection of accented Bengali language (Chittagong Accent). We collected a total of 3161 voice data from 25 classes. We collected the audio data in the following way:

1. **Manually:** We had a team to collect the audio data from the local people of Cox's Bazar. We did this manually and tried to use the same type of microphone to maintain a specific Hz frequency of the audio.
2. **Normalization:** We converted all the audio file into WAV format and the same sample rate using a software called Audacity. The used audio sample rate was 1600 kHz.
3. **Segmentation:** Each audio file consists of a speaker saying each words 5 times of total 25 words. We segmented each words audio separately and stored them separately.
4. **Conversion:** One of our research methods needed audio Spectrograms of training Neural network model. That's why we needed to convert the audio files into Mel Spectrograms and store them secretly. For this issue, we used a python library called 'Librosa' to convert the audio files to mel Spectrograms. Usually while converting audio to Mel Spectrogram, we need to convert all the audio file into a same number of channels. An audio file can either consist of two or one channel. Stereo or Mono channel or both. Our audio file was mixed. Some had stereo some hand mono, and some had both. So, converting each 3161 audio files into a single channel was a big hassle. That's why we used librosa for this purpose. Librosa doesn't require the same number of channels for each audio

files it converts it automatically. That's why we didn't need to write any extra code or do it manually using Audacity to convert them into a single channel.

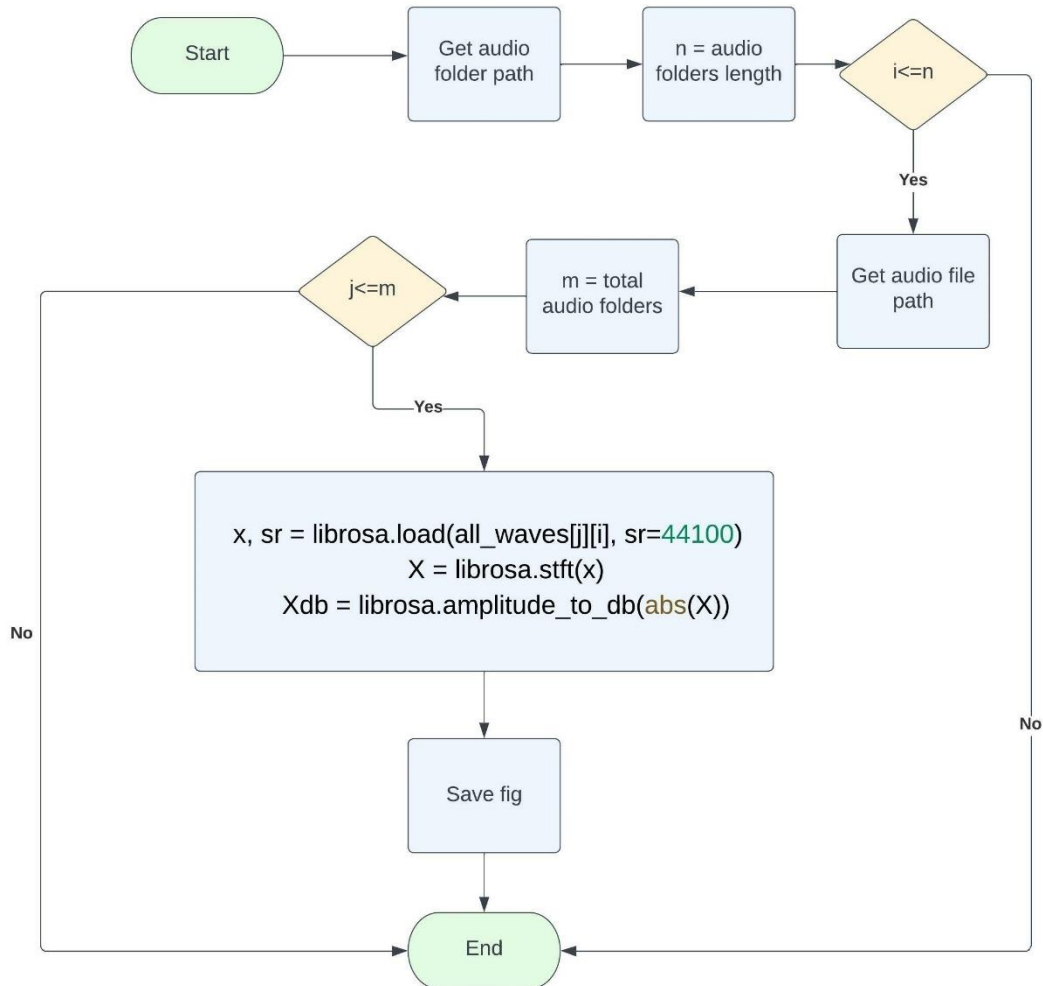


Figure 3.1: Flowchart of audio to spectrogram conversion algorithm

3.2 Dataset utilized

In this research we used audio data which were on average of 3-4 minutes and were collected from 40-50 people. Speakers were aged of 20-70 years. The audio file mainly contained 25 words or classes. After the data split, we had in total of 3161 audio files of 25 classes or different words. As we all know we cannot straight up use the audio file to train the machine learning or deep learning models. For this reason, we need any feature extractor to extract the feature from the audio files of a format that can be used

to train the machine learning or deep learning models. For this reason, we used two methods to extract features from the audio files we collected.

1. Use of a feature extractor ex. Mfcc

2. Image approach ex. Converting audio to Mel spectrograms.

This section will provide the detail of the above-mentioned techniques.

- 1. MFCC feature extractor:** Mfccc are commonly used to extract features from audio signal so that a speech detection system can be trained and tested with it. The steps of mfcc feature extraction are as follows: windowing the signal, applying discrete Fourier transform on it, log of the magnitude, wrapping the frequency on mel scale, take logarithm of it, inverse DFT, mel cepstrum and then mfcc features vector:

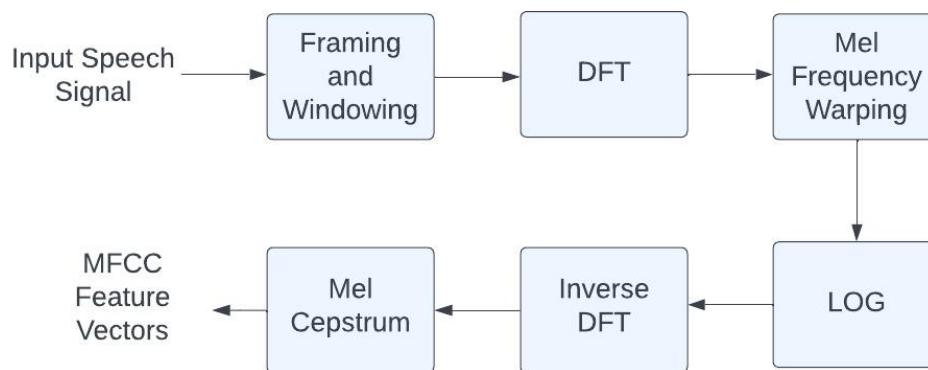


Figure 3.2: Steps of generating mfcc

1.1. Windowing/Framing the signal:

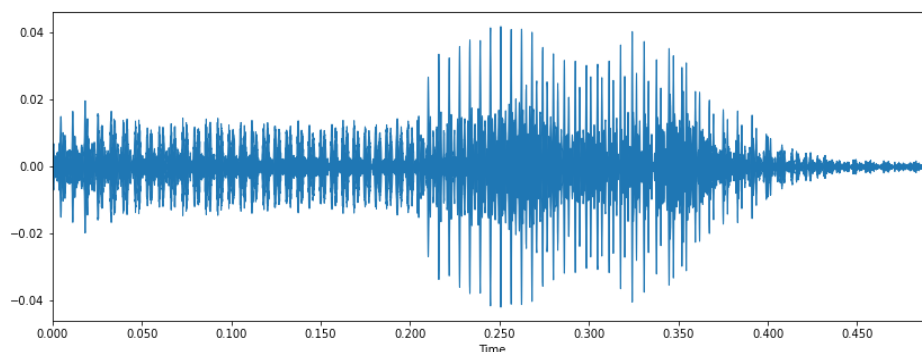


Figure 3.3: Wave Plot of Audio Signal

Windowing the signal mean window, it on temporal dimension and apply Fourier transformation on each window instead of applying it on the whole signal at once. We do that because applying Fourier transformation is not very informative to the change's nature of the audio. This could lead to loss of frequency over time. The standard windowing interval is between 20-40ms where 25ms is the standard and in 10 steps. The interval between one window to another is 10ms.

1.2 Cepstrum:

$$C_p = |F\{\log(|F\{f(t)\}|^2)\}|^2 \dots \dots (1)$$

Cepstrum is computed as the above equation. The steps are: DFT => Mel frequency wrapping => Log => Invers DFT or DFT. First, we apply Fourier transform $|F\{f(t)\}|$. Then we apply logarithm of it then we do the Fourier transform on it again. The last step also can be inverse Fourier transform but the output of both is same. They are almost equivalent only differ in a magnitude factor. The cepstrum is an anagram of a spectrum.

1.3: Converting into Mel scale: This is the first step we need to do before doing step 1.1 and 1.2.

$$f = 700 \left(10^{\frac{m}{2595}} - 1 \right) \dots \dots \dots (2)$$

Converting an audio signal into Mel scale means we can get more information about the human auditory system. This means we mainly focus on the part that human listeners find important in an audio signal. For that first, we need to take k points on the Mel frequency axis and compute the k frequency against it on the frequency axis. We compute it using the formula (2). But only working with important frequencies can throw away important energy information.

$$u_k = \sum_{h=f_{k-1}+1}^{f_{k+1}-1} w_{k,h} |x_h|^2 \dots \dots \dots (3)$$

This issue is fixed by taking weighted average around this frequencies. This technique is known as triangular filter bank. We compute it with equation (3) where $w_{k,h}$ is are the points that we get from DFT, x_h are the energy value corresponding to frequency domain.

2. Mel Spectrogram: As we mentioned before mfcc is computed by performing log and DFT on Mel spectrogram. MFCC are more compressed. It has 20 or 13 coefficients on the other hand mel spectrograms works with 32 to 64 bands. This means MFCC features are more decorated that's why works well with linear machine learning models like gaussian mixture models on the other hand mel spectrograms performance is better with strong classifiers that works with huge data's ex. ANN, CNN. In the mel spectrogram the audio which are in Hz are converted into mel scale. The spectrograms work well when all the frequencies are equally important on the other hand mel spectrograms works well when model needs human auditory perception.

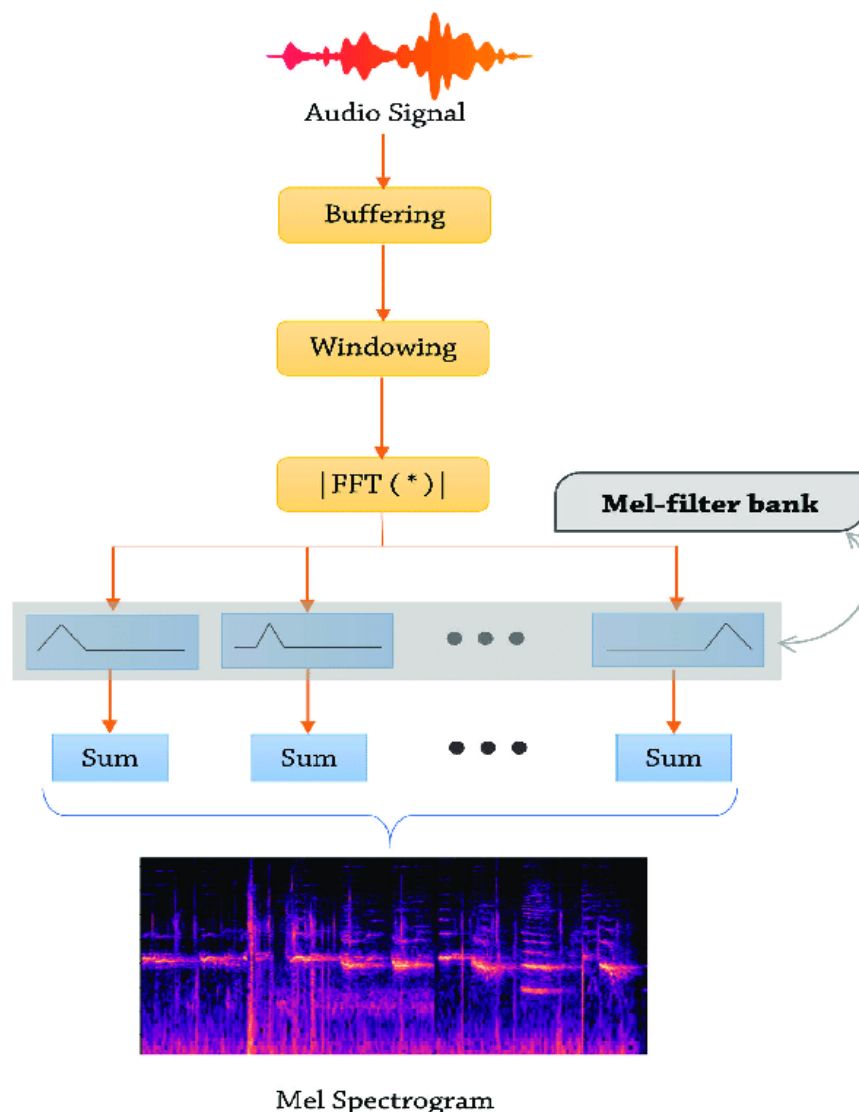


Figure 3.4: Generation of Mel Spectrogram

The first step is windowing the audio signal as described in section 1.1. Then STFT or short time Fourier transform equation (4) is applied to each window.

$$STFT(t, w) = \int_{-\infty}^{\infty} f(\tau)W(\tau - t)e^{-i\omega\tau}d\tau \dots \dots \dots (4)$$

Then Mel filter bank is applied. Mel filter bank basically decompose the audio into a different frequency which actually mimics the nonlinear human auditory perception of the audio signal. After, that we apply logarithm on the sum of it and generate Mel spectrogram from it.

3.3 Statistical Analysis

In this section, we will provide further details about the dataset. We previously mentioned the amount of data and the number of classes in the dataset. Table 3.1 includes additional information about the dataset.

Table 3.1: INFORMATION OF DATASET

Class	Represented number of	Total data
আই	0	140
আঁর	1	126
আছন	2	130
ইবা	3	126
উড়াউম	4	137
উয়ারে	5	125
একডইল্লা	6	131
গুগগা	7	126
কিয়াল্লায়	8	126
কেন	9	135
কেনগরি	10	126
গরি	11	121
ইন্তে	12	112
হডে	13	100
হতা	14	122
হন	15	131
হনডে	16	140
হাঁদা	17	123

হাওয়া	18	132
হাছহাছি	19	122
হারন	20	125
হারাপ	21	131
হারে	22	145
হালুয়া	23	129
হোনা	24	131

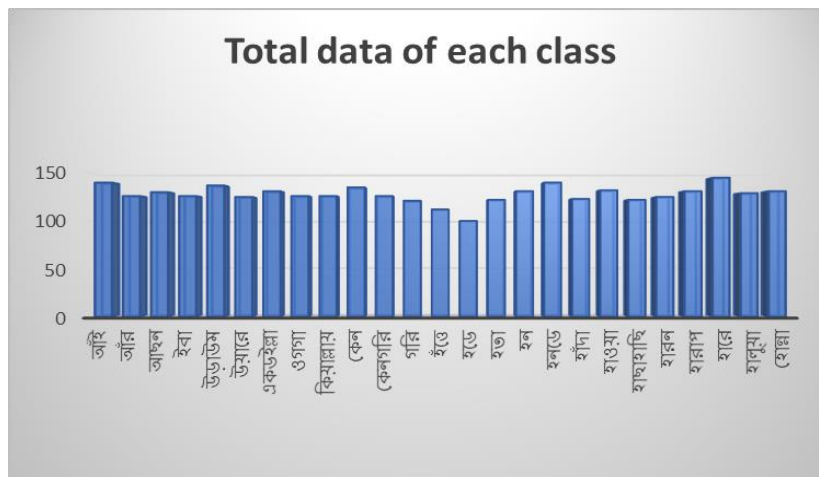


Figure 3.5: Total Data of Each classes

We Collect every word as in same amount but for some noise of we had delete some of the data. That's why in every class there are not same amount of data.

3.4 Methodology

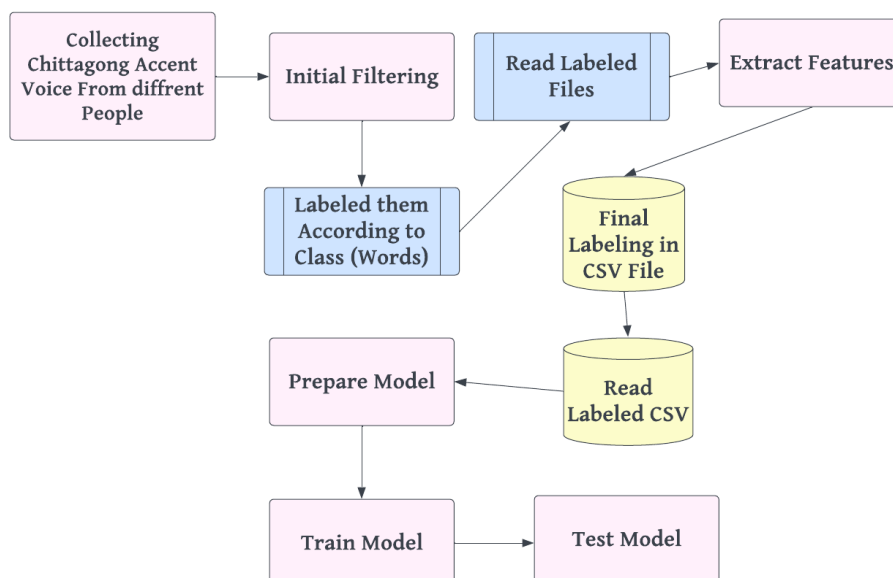


Figure 3.6: Methodology Figure

As the Figure shows first, we collected audio. Then we split the audio as 25 classes. Then saved and stored it in a folder structure with their class label. After that we extracted features from the audio files and trained model with those data.

3.5 Implementation Requirements

Python 3.11.0:

Python is a high-level programming language that is interpreted, object-oriented, and has dynamic semantics. [7] High-level in-built data structures, along with dynamic type and binding, make this language appealing for use in Rapid Application Development, as well as for use as a scripting or glue language to tie together pre-existing components. The low cost of Python's upkeep can be attributed to the language's focus on readability. Python's module and package system facilitates and promotes program modularity and the reusability of code. There is no cost to share either the Python interpreter or the vast standard library, both of which are freely accessible in source and binary form for all major platforms. That makes it much simpler for us to get our jobs done. Python's widespread popularity stems in large part from the fact that it is both simple and easy to learn. In addition, the most recent version, 3.11.0, was employed in our studies.

Anaconda 2022.2.1:

For scientific computing, such as data science, machine learning applications, large-scale data processing, predictive analytics, etc. [8] Anaconda is a distribution of the Python and R programming languages that attempts to ease package management and deployment. [8] This is actually an installer for a package. By installing only one thing, it will install a large number of other data science tools that are important. Even so, it is packaged along with the idea of a virtual world. We may keep distinct projects apart from one another so that we can tailor the criteria that we utilize for each project specifically. We utilized version 2022.2.1 of anaconda, which was the most recent version available at the time.

Jupyter notebook:

To create and share computational documents online, the Jupyter Notebook was the first web application to do so. It provides a user-friendly, simplified, and document-

focused environment. [9] When working on a Jupyter project, the Jupyter Code of Conduct applies to all of your conversations, both online and off, with other members of the team. Our goal in creating this Code of Conduct is to ensure that everyone involved in this project is treated with dignity and respect. [9] To be more precise, it's a web-based open source that enables the aforementioned as well as much more, such as the creation of code, the visualization of data, the use of equations, and much more. Jupyter Notebook version 6.0.3 was used for this project.

Keras:

Google's Keras provides a high-level API for deep learning and the use of neural networks. [10]. It simplifies the process of putting together neural networks, and it's built in Python. It can also compute using many neural networks in the background. This library simplifies the process significantly. They have already constructed the calculating components, so all we have to do is plug them in to see where they go to meet our requirements. TensorFlow is used as the backend for Keras. There are other libraries that use Keras as a backend, but Keras is the most mature and feature-rich of the bunch. Keras 2.11.0 and TensorFlow 2.11.0 were the libraries of choice. The deep neural network was tested using Keras and TensorFlow in this study.

Scikit learn:

This is a free library written in Python that contains a variety of algorithms for classification, clustering, and regression [11]. The implementation of such algorithms is simplified as a result. While working on this project, we pulled various algorithms from the library and used them. The following is a list of the algorithms that were used:

- ❖ SVM-based support vector machine
- ❖ The K-nearest neighbor method
- ❖ Logistic regression
- ❖ Random forest
- ❖ Gradient boosting

Pydub:

Pydub is a Python library that does not cost anything [18]. Typically utilized for straightforward audio processing. The following is a list of some of the functions that pydub offers:

- ❖ Read audio files
- ❖ Play audio files
- ❖ Split audio files
- ❖ Audio files and a whole lot more may be found in Marge.

Librosa:

Librosa is a Python tool that may be used to analyze music and audio. [19]. When we work with audio data, such as when generating music (using LSTMs) or doing automatic speech recognition, Librosa is the primary tool that we employ. It offers the foundational components that are essential for the development of music information retrieval systems. Because analog data cannot be processed by computers, it is necessary for us to turn it into numeric values. Librosa was the one that completed this task for us. We started with a single audio file and retrieved 5 characteristics from it. The version of librosa that we used was 0.9.2.

Audacity:

Audio editing and recording for Windows, macOS, GNU/Linux, and other platforms have never been easier than with Audacity. The audio editing program Audacity may be downloaded for free and legally. [20]. Specifically, we put it to use by altering the files' aural characteristics. We used Audacity to cut up a large file into smaller pieces, merge them back together, and then normalize the volume by converting to 16 kHz. The Audacity 3.2.3 version was used.

Chapter 4

Experimental Result and Discussion

4.1: Experimental setup

On chapter 3 we showed how we collected the audio data and preprocessed them using audacity and extracted features from them by converting them to Mel spectrograms and use of Mfcc feature extractor. In this chapter, we will see how we used these features to train our model and reach our goal. We used two machine learning algorithms Gradient boosting classifier and support vector machine and two deep learning algorithms convolutional neural network CNN and artificial neural network ANN. We used the Mfcc features that we extracted from audio data to train the Gradient boosting classifier, Support vector machine, Artificial neural network ANN. On the other hand, we used an image classifier approach to train CNN or convolutional neural network model. This image classifier approach was mainly converting the audio signals into Mel spectrograms.

We had in total of 3161 audio files of 25 classes. After converting them into Mel spectrograms we used them to train out CNN model. We used 80% data for training the CNN model and 20% data for testing and validation of the model. The CNN model has two convolutional layers, one fully connected layer, one Softmax layer.

Testing the model with 20% of test data we get the result shown in the below confusion matrix and train accuracy vs validation accuracy and train error and validation error curve. Figure 4.1 is the normalized confusion matrix of Convolutional neural network model. It shows the performance of each 25-word classes. Figure 4.2 and Figure 4.3 shows the train accuracy and validation accuracy, and train loss and validation loss curves are shown. Other three models' figure are also provided below in this sequence. Figure 4.4 shows a full picture of how the CNN model performed for each classes.

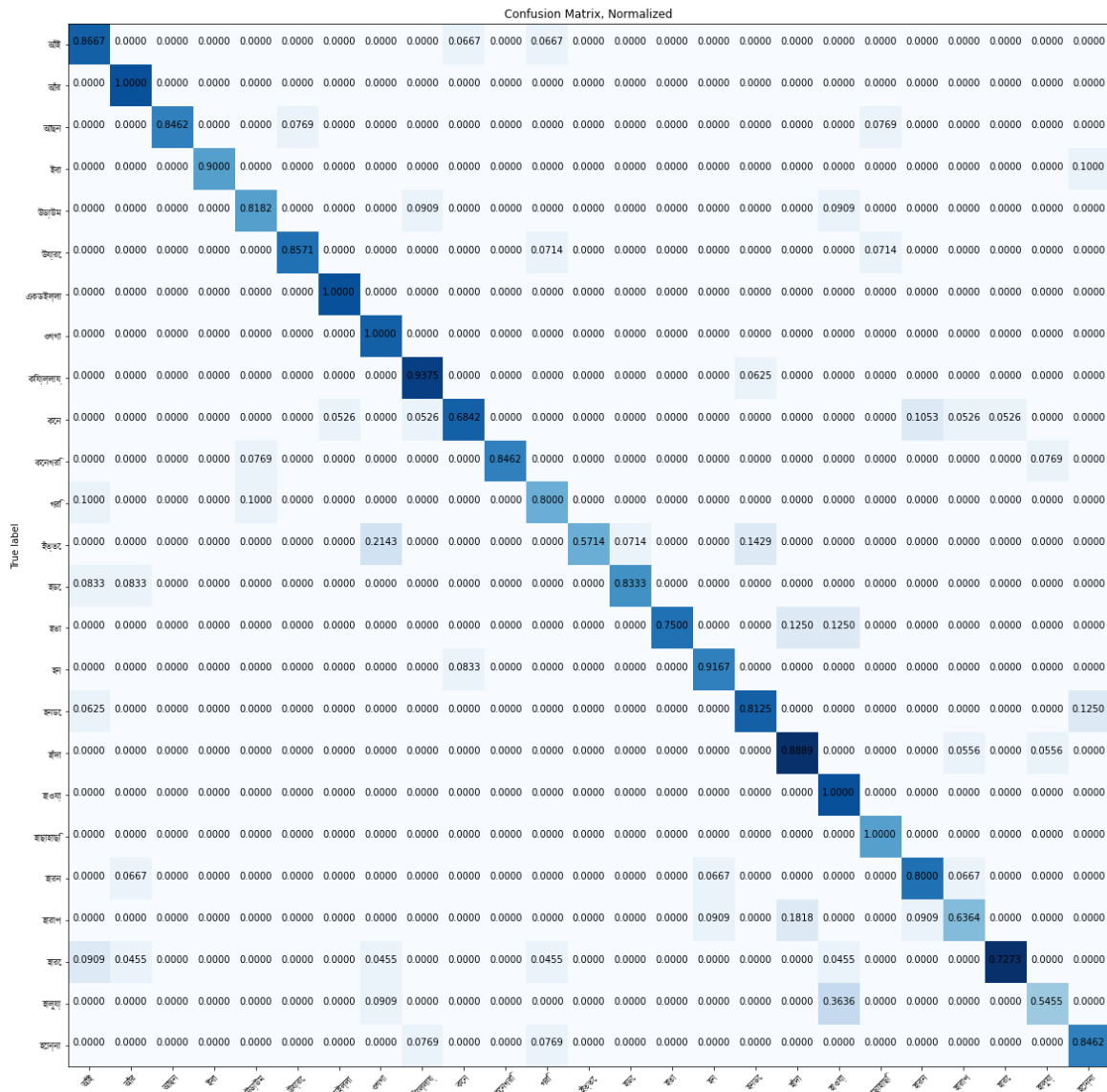


Figure 4.1: Normalized confusion matrix of CNN

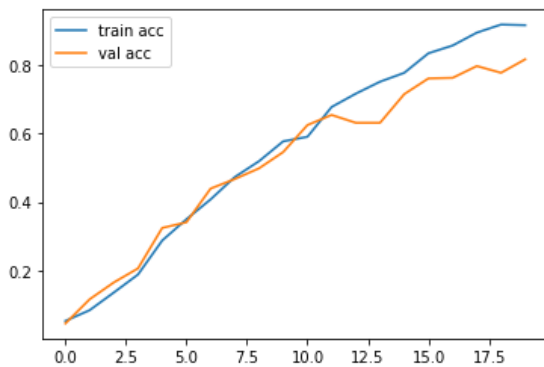


Figure 4.2: Train accuracy vs Validation accuracy curve

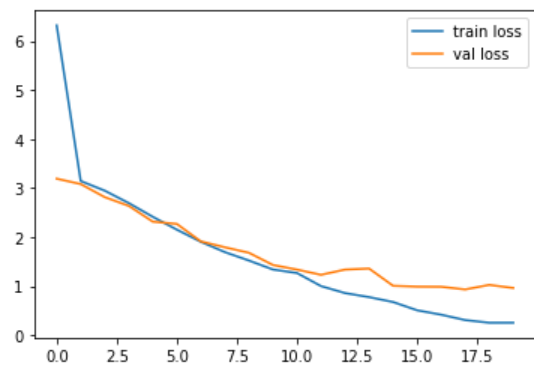


Figure 4.3: Train loss and Validation loss

A machine learning model's performance may be inferred from the graph of train accuracy and validation accuracy against the number of epochs (x-axis), as shown in

Figs. 4.2 and 4.3. When both the train accuracy and validation accuracy rise as the number of epochs rises, the model is likely learning and becoming more accurate. The stopping conditions, such as the maximum number of epochs or a minimal change in accuracy, will determine the final accuracy score.

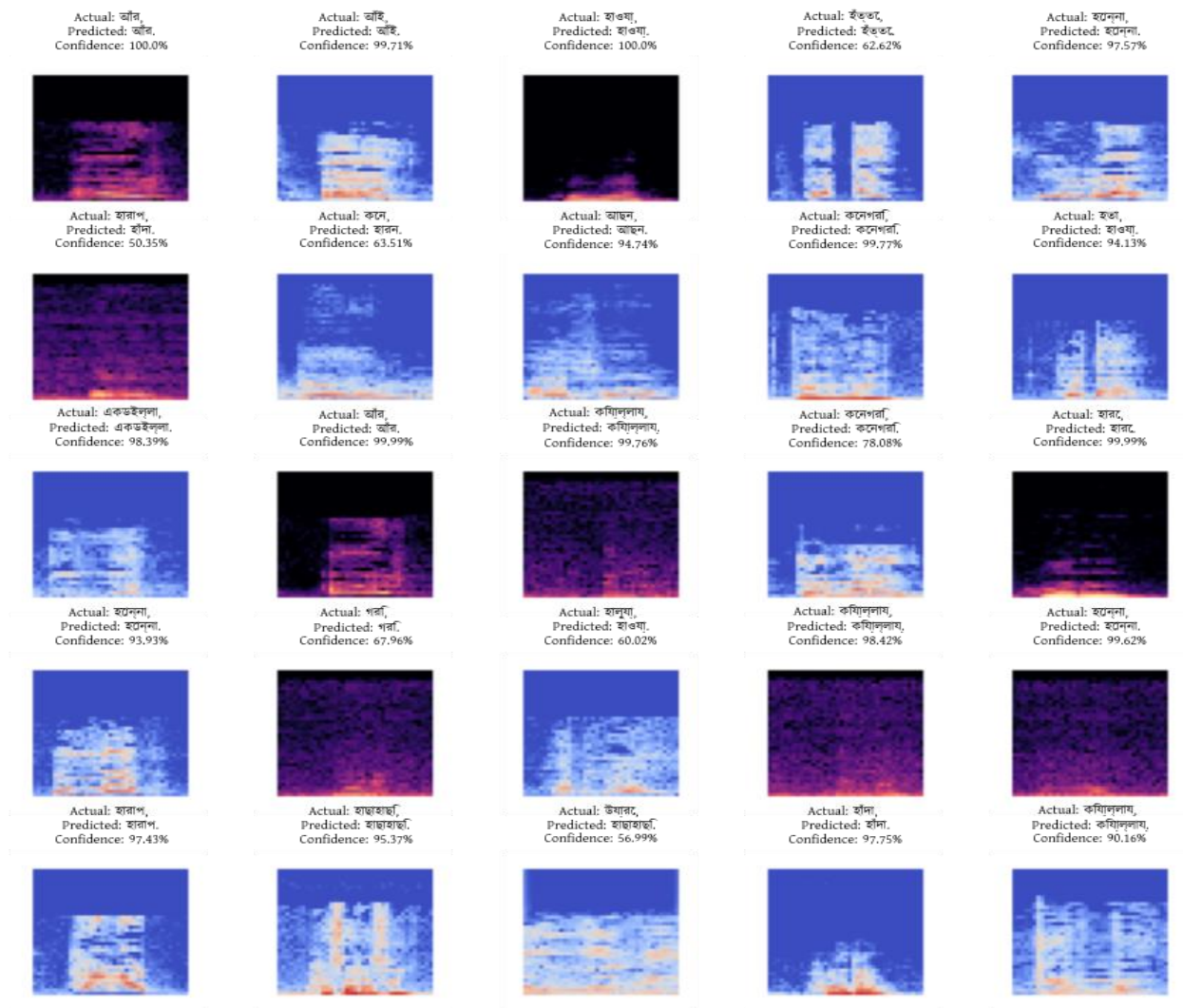


Figure 4.4: Actual and predicted classes and confidence percentage of CNN model

We also trained two machine learning models Gradient boosting classifier and Random Forest Classifier algorithm. For feature extraction from audio data, we used MFCC with it. Generally, MFCC performs better with these linear machine learning algorithms. Below are two confusion matrix those interprets the model's performance with 20% test data.

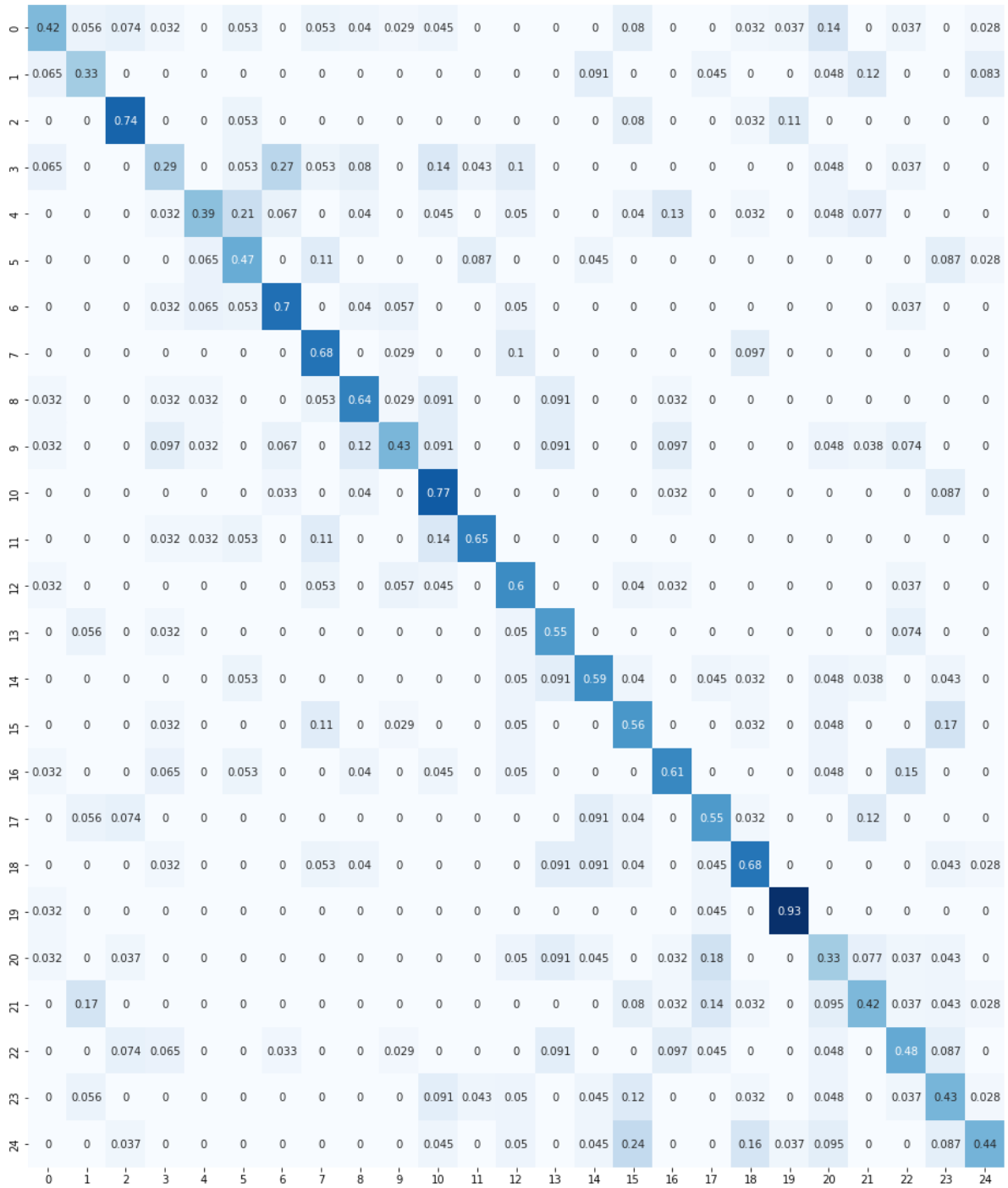


Figure 4.5: Confusion matrix of Gradient boosting classifier model

0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24 corresponds to the class labels 'আই', 'আঁ', 'আছন', 'ইবা', 'উড়াউম', 'উয়ারে', 'একডইল্লা', 'ওগগা', 'কিয়াল্লায়', 'কেন', 'কেনগরি', 'গরি', 'ইন্তে', 'হডে', 'হতা', 'হন', 'হনডে', 'হাঁদা', 'হাওয়া', 'হাছাছাছি', 'হরন', 'হরাপ', 'হরে', 'হলুয়া', 'হোনা'

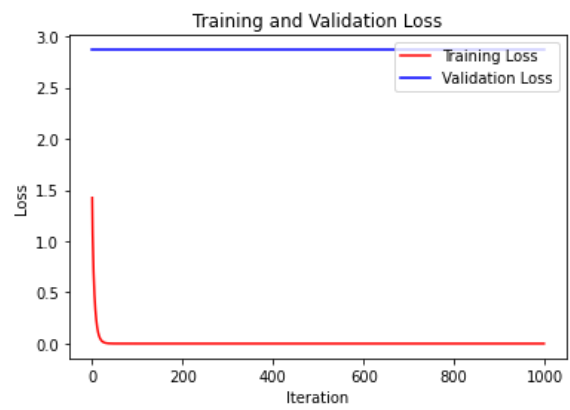
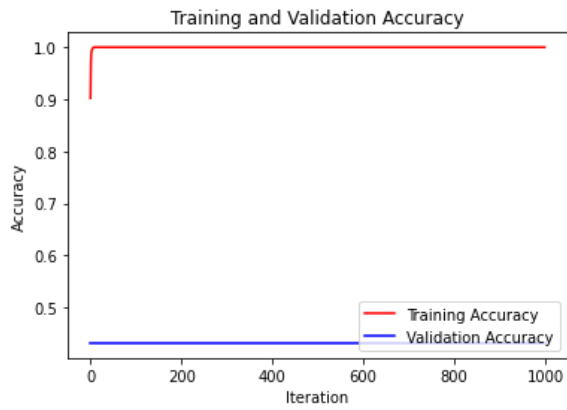


Figure 4.6: Train accuracy vs Validation accuracy curve

Figure 4.7: Train loss and Validation loss

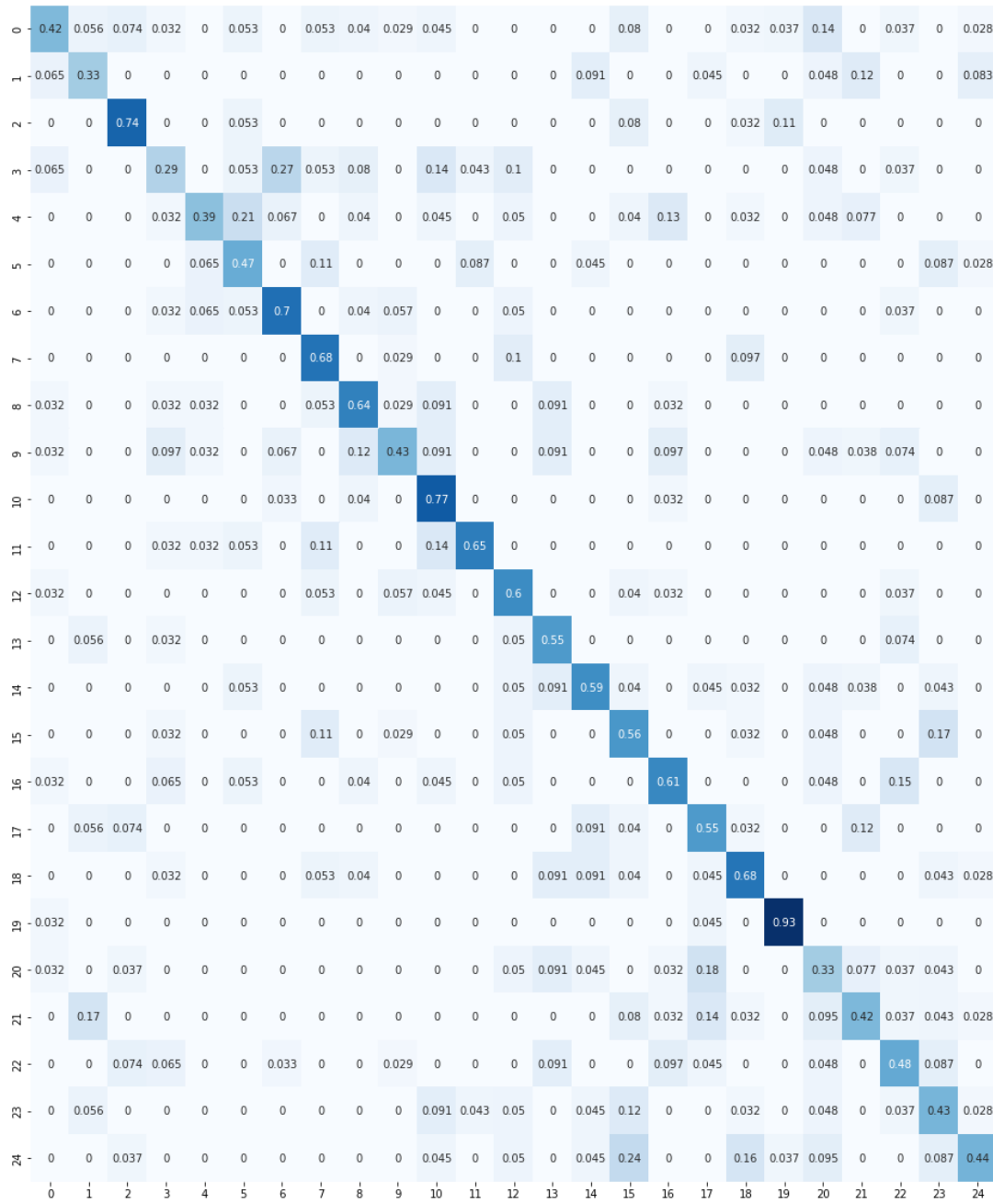


Figure 4.8: Normalized Confusion matrix of Random Forest classifier



Figure 4.9: Train accuracy vs Validation accuracy curve



Figure 4.10: Train loss and Validation loss curve

When the train accuracy and validation accuracy curves in fig. 4.6, 4.7, 4.9, and 4.10 move in parallel as the number of epochs grows along the x-axis, it means that the model is probably generalizing well to new, unknown data. This indicates that the model is capable of maintaining its performance on both the training and validation sets and is not overfitting to the training data.

The stopping conditions, such as the maximum number of epochs or a minimal change in accuracy, will determine the final accuracy results. A model is generally considered to be well-trained and capable of generalization to new data if it performs well on both the training and validation sets.

Artificial neural network (ANN) was a different model we employed. In contrast to CNN, MFCC was used in this instance as the feature extraction approach. The confusion matrix that analyzes the performance of the model using 20% test data is shown below.

0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24 corresponds to the class labels 'আই', 'আঁর', 'আছন', 'ইবা', 'উড়াউম', 'উয়ারে', 'একডইল্লা', 'ওগগা', 'কিয়াল্লায়', 'কেন', 'কেনগরি', 'গরি', 'ইত্তে', 'হডে', 'হতা', 'হন', 'হনডে', 'হাঁদা', 'হাওয়া', 'হাছাহাছি', 'হারন', 'হারাপ', 'হারে', 'হালুয়া', 'হোনা'

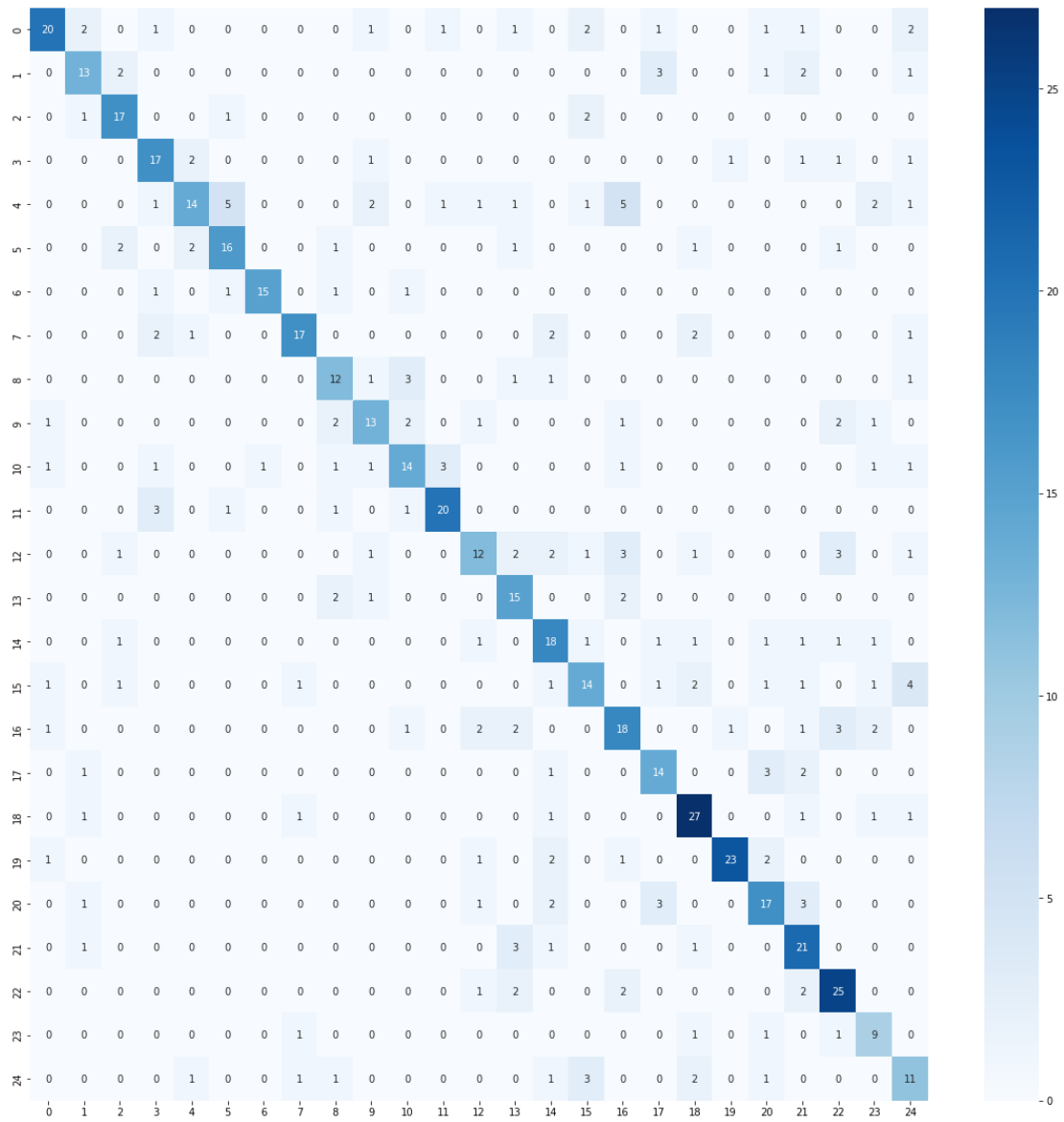


Figure 4.7: Confusion Matrix of ANN

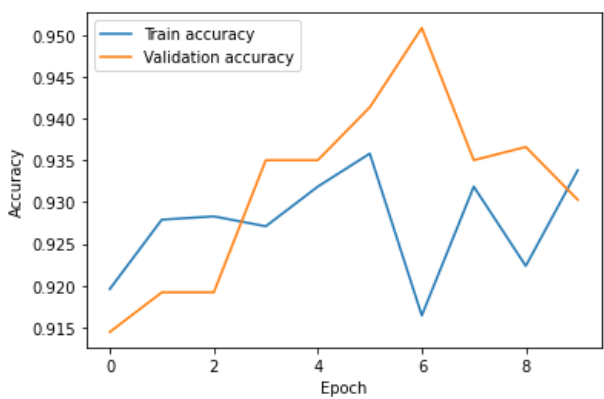


Figure 4.8: train accuracy vs validation accuracy curve



Figure 4.9: train loss and validation loss

If the train accuracy and validation accuracy curves in figs. 4.8 and 4.9 plateau or decline after a given number of epochs along the x-axis, it may mean that the model

has hit its maximum performance on the data and that more training will not increase accuracy. In certain circumstances, continuing to train the model after this stage may result in overfitting, when the model starts to remember the training data and performs badly on new, unforeseen data (validation set).

In other instances, the model may have simply learned all that it could from the data, and more training will not increase accuracy. In either scenario, the model may benefit from additional model architecture fine-tuning or from gathering more varied and representative training data.

The model had three hidden layer and input output layers. Batch normalization was done in each of the layers except the output layer. Activation function is ‘eli’. Below a total summary of the model given.

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 64)	2624
activation (Activation)	(None, 64)	0
batch_normalization (Batch Normalization)	(None, 64)	256
dense_1 (Dense)	(None, 64)	4160
activation_1 (Activation)	(None, 64)	0
batch_normalization_1 (Batch Normalization)	(None, 64)	256
dense_2 (Dense)	(None, 64)	4160
activation_2 (Activation)	(None, 64)	0
batch_normalization_2 (Batch Normalization)	(None, 64)	256
dense_3 (Dense)	(None, 25)	1625
activation_3 (Activation)	(None, 25)	0
=====		
Total params: 13,337		
Trainable params: 12,953		
Non-trainable params: 384		

Figure 4.10: ANN model summary

4.2: Experimental Result and analysis

We finished data collection, preprocessing, feature extraction, and methodology setting in the previous chapter. We need to use different algorithms in accordance with the current technique, and CNN network produces superior results, so we'll talk about these three procedures in this section.

Among the four algorithm CNN performed best in our dataset. In table 4.1 we get a comparative performance of the four used algorithms.

Table 4.1: Comparison table of 4 classifiers

Model	Val accuracy	Train accuracy	Error rate	Precision	Re-call	F1-score
Convolutional neural network (CNN) with Mel spectrograms	81%	90%	13	84	84	83
Gradient boosting classifier with MFCC features as input	44%	56%	55	43	44	43
Random forest classifier with MFCC features as input	62%	70%	37	62	62	61
Artificial neural network with MFCC	73%	81%	26	76	69	72

As we can see CNN performed best with 81% validation accuracy and over 90% training accuracy. Generally, MFCC features are more decorated that's why works well with linear machine learning models like gaussian mixture models on the other hand

mel spectrograms performance is better with strong classifiers that works with huge data. Though the ANN model seems performed good with MFCC features but in one cost. The initial performance of the ANN model was very low. The train accuracy was around 40%.

To increase the performance, we made architecture of the model more complex added more hidden layers and batch normalization. Though it gave higher accuracy but as we can see in Figure 4.9 there is a huge gap between train loss and validation loss curve. This means in order to get better performance the model got overfitted. On the other hand, though some audio files had noise in it converting those into Mel spectrograms gave a good performance with strong classifier like CNN. When employing MFCCs for noisy audio data, there are a number of issues that could result in an ANN model having low accuracy. The following are a few potential causes.: Insufficient training data: Without adequate training data, the model may not be able to learn well or generalize to new samples. Poor data quality: It may be challenging for the model to extract usable characteristics from audio data that is distorted or of poor quality. Inappropriate model architecture: A model architecture that isn't suitable for the task or the data will perform poorly. Inadequate training: The model may not converge to a good solution if it is not taught for a sufficient number of epochs or if the learning rate is not set correctly. You may need to attempt several methods, like as gathering more data of higher quality, experimenting with various model architectures, or modifying the training hyperparameters, in order to increase the model's accuracy. To increase the signal-to-noise ratio of the audio data, it may also be beneficial to experiment with various feature extraction techniques or include additional preprocessing stages like denoising or spectral subtraction. It is not surprising that a CNN model might outperform MFCCs for noisy audio data when Mel spectrograms are used instead. This is because to CNNs' capacity to learn spatial correlations in the data, which may allow them to utilize the additional spectral data offered by Mel spectrograms more effectively. However, the particular qualities of the data and the task at hand will determine the choice of feature representation. To determine which feature representation is most effective for your specific application, you may want to experiment with both MFCCs and Mel spectro

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1: Impact on Society

The application of natural language processing (NLP) to the research of accent identification, particularly in the Chittagong region of Bangladesh, has the potential to have a significant influence in this area. Researchers can examine the Chittagong accent and its differences from other accents by employing natural language processing tools. Applications include research into the sociocultural elements that shape language usage in the Chittagong region and the development of voice recognition systems that are better equipped to grasp and transcribe spoken language in the region. Further, natural language processing (NLP) can assist researchers in recognizing patterns and trends in the usage of the Chittagong accent, therefore revealing interesting information about the linguistic practices of the inhabitants of this area. We can learn a lot more about language and its function in society if we put NLP to work in accent detection studies.

5.2: Impact on Environment

The effects of Bangla NLP, especially in the Chittagong dialect, on the natural world can be beneficial. This software's many potential uses include facilitating the creation of automated systems for translating environmental legislation and directives into Bangla, thus increasing their reach. Policymakers and environmental groups may utilize this data to better understand and respond to community needs, resulting in less paper and other physical materials being consumed in the communication and information distribution processes.

5.3: Ethical Aspects

Our information has not been taken from another website on the internet. We go through each person who is originally from Chittagong and get raw data from them. We take precautions to ensure the safety of our data so that it is not taken by unauthorized

parties. We are sensitive to the need to protect the personal information of persons whose data we collect.

5.4: Sustainability Plan

A sustainability plan is a plan that specifies how a company, organization, or community will function in a way that is sustainable economically, socially, and ecologically. Our application will be designed for end-to-end users, and we will continue to improve it in order to provide our clients with access to the most recent features as they become available.

CHAPTER 6

Summary, Conclusion, Recommendation and Implication for Future Research

6.1 Summary of the Study

The purpose of this research project is to predict the accents of different regions in Bangladesh. This is important in order to maximize the use of the automatic speech recognition (ASR) system in all parts of Bangladesh. Additionally, there are many practical applications for this technology. In order to achieve this goal, we utilized the Mel-frequency cepstral coefficients (MFCCs) technique to extract features from speech recordings and employed various machine learning and deep learning techniques to train the system. We divided our dataset into an 80% training set and a 20% testing set. The training set was used to teach the system, while the testing set was used to evaluate its accuracy.

6.2 Conclusion

The focus of this research is the development of a speech recognition system for Bengali, a language spoken in Bangladesh. The number of people who speak Bangla worldwide is staggering [12]. Bangla is the native tongue of Bangladesh and is also spoken in some regions of India [13]. One of the challenges in developing such a system is that the standard Bengali language is not spoken uniformly across the country and there are many local accents. To address this issue, the researchers used a combination of techniques including a convolutional neural network to convert audio into a spectrogram, gradient boosting with Mel Frequency Cepstral Coefficients (MFCC) features of audio data, and a Random Forest classifier. The system was trained and tested using audio files from the local accent of the people of Chittagong and achieved a training accuracy of 91% and a test accuracy of 81% with 3161 audio files in 25 classes. The researchers hope to continue their work on this topic in order to make the system more widely applicable to speakers of various local accents.

6.3 Implication for Further Study

It is important to always strive for improvement and make things better. This is also applicable in the context of this work, as there are likely many aspects that can be enhanced. Some potential areas for future improvement are listed below:

- 1.** In the Bangla language, there is a large number of accents that we have only worked with a small portion of. By increasing the number of accents, we utilize, we have the opportunity to expand our linguistic abilities and better understand the diversity and complexity of the Bangla language.
- 2.** In order to make our model stronger, we need to incorporate more diversity in our speaker selection. A variety of perspectives and experiences can enrich the discussions and presentations, leading to a more well-rounded and robust model. By including a range of speakers with different backgrounds, we can ensure that all viewpoints are represented and that our model is stronger as a result. It is important to recognize the value of diversity and to actively seek out and include a variety of speakers in our model.
- 3.** We are experimenting with various classifiers in order to find the most effective one for our needs. This involves testing multiple options and comparing their performance in order to determine which is the best fit for our specific situation. By trying out different classifiers and comparing them, we hope to find the one that will most accurately classify the data we are working with. We believe that this process of trying multiple options and comparing their results is important in order to ensure that we are using the most suitable classifier for our purposes.

REFERENCES

- [1] Nguyen and Q. Trung, "Speech classification using SIFT features on spectrogram images," *Vietnam Journal of Computer Science*, vol. 3, no. 4, pp. 247-257, November 2016.
- [2] P. D. and M. Woźniak, "Image approach to voice recognition," *IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1-7, November 2017.
- [3] K. Kubicki, P. Kapusta and K. Ślot, "Presentation Attack Detection on Limited-Resource Devices Using Deep Neural Classifiers Trained on Consistent Spectrogram Fragments.," *Sensors*, vol. 21, no. 22, p. 7728, November 2021.
- [4] A.-H. O. . A. Mohamed, H. Jiang, L. Deng, G. Penn and D. Yu, "Convolutional Neural Networks for Speech Recognition," *IEEE/ACM Transactions on audio, speech, and language processing*, no. 10, pp. 1533-1545, July 2014
- [5] Badshah, A. J. Ahmad, N. Rahim and S. W. Baik, "Speech emotion recognition from spectrograms with deep convolutional neural network," *2017 international conference on platform technology and service (PlatCon)*, pp. 1-5, February 2017.
- [6] Al-Darkazali and M, "Image processing methods to segment speech spectrograms for word level recognition.," *Doctoral dissertation, University of Sussex*, 2017.
- [7] Yuan, L and . J. Cao, "Patients' EEG data analysis via spectrogram image with a convolution neural network," *International Conference on Intelligent Decision Technologies*, pp. 13-21, June 2017.
- [8] Kalimullah, K. and Sushmitha, D, "Influence of design elements in mobile applications on user experience of elderly people." *"Procedia computer science"*, 113, pp.352-359, 2017.
- [9] Dennis, Jonathan, H. D. Tran and L. Haizhou, "Spectrogram image feature for sound event classification in mismatched conditions," *IEEE signal processing letters*, pp. 130-133, December 2010.
- [10] Salau, A.O, T. Olowoyo and S. Akinola, "Accent classification of the three major nigerian indigenous languages using 1d cnn lstm network model," *In Advances in Computational Intelligence Techniques*, pp. 1-16, 2020.
- [11] Georgina Brown, Automatic Accent Recognition Systems and the Effects of Data on Performance, Odyssey June 2016, pp.21-24, Bilbao, Spain
- [12] "Bengali language, 2017. [Online]. Available: https://en.wikipedia.org/wiki/Bengali_language [Last accessed: 1- Jan- 2023]
- [13] Mamun, R.K., Abujar, S., Islam, R., Badruzzaman, K.B.M. and Hasan, M. "Bangla speaker accent variation detection by MFCC using recurrent neural network algorithm: a distinct approach." In *Innovations in Computer Science and Engineering*. pp. 545-553. Springer, Singapore, 2020.
- [14] "Bengali language", 2023. [Online]. Available: https://en.wikipedia.org/wiki/Bengali_language [Last accessed: 1- Jan- 2023]
- [15] "Chittagong language", 2023. [Online]. Available: https://en.wikipedia.org/wiki/Chittagonian_language [Last accessed: 1- Jan- 2023]

- [16]“Alexa” [Online]. Available: <https://developer.amazon.com/en-US/alexa> [Last accessed: 1-Jan- 2023]
- [17]“Siri” [Online]. Available: <https://www.apple.com/siri/> [Last accessed: 1- Jan- 2023]
- [18]“Pydub” [Online]. Available: <https://pydub.com>[Last accessed: 1- Jan- 2023]
- [19]“Librosa” [Online]. Available: <https://librosa.github.io/librosa/> [Last accessed: 1- Jan- 2023]
- [20]“Audacity” [Online]. Available: <https://www.audacityteam.org/> [Last accessed: 1- Jan- 2023]

Project_Report

ORIGINALITY REPORT

11%

SIMILARITY INDEX

7%

INTERNET SOURCES

2%

PUBLICATIONS

6%

STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	3%
2	Muhammadjon Musaev, Ilyos Khujayorov, Mannon Ochilov. "Image Approach to Speech Recognition on CNN", Proceedings of the 2019 3rd International Symposium on Computer Science and Intelligent Control, 2019 Publication	1%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
4	Submitted to University of Strathclyde Student Paper	1%
5	en.wikipedia.org Internet Source	1%
6	Submitted to University of Northumbria at Newcastle Student Paper	1%
7	www.ijraset.com Internet Source	<1%

8	myassignmenthelp.com Internet Source	<1 %
9	Submitted to University of College Cork Student Paper	<1 %
10	Submitted to Muscat College Student Paper	<1 %
11	Submitted to ABES Engineering College Student Paper	<1 %
12	Submitted to Coventry University Student Paper	<1 %
13	Submitted to University of Greenwich Student Paper	<1 %
14	penerbit.uthm.edu.my Internet Source	<1 %
15	ijcrt.org Internet Source	<1 %
16	Yann Bayle, Matthias Robine, Pierre Hanna. "SATIN: a persistent musical database for music information retrieval and a supporting deep learning experiment on song instrumental classification", Multimedia Tools and Applications, 2018 Publication	<1 %
17	dev.healthmanagement.org Internet Source	<1 %

18	www.slideshare.net Internet Source	<1 %
19	1library.net Internet Source	<1 %
20	Christopher B. Anderson. "The CCB-ID approach to tree species mapping with airborne imaging spectroscopy", PeerJ, 2018 Publication	<1 %
21	citeseerx.ist.psu.edu Internet Source	<1 %
22	oatao.univ-toulouse.fr Internet Source	<1 %
23	www.mdpi.com Internet Source	<1 %
24	www.sciencegate.app Internet Source	<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography On