

**HUMAN SPEECH EMOTION CLASSIFICATION BASED ON BENGALI
LANGUAGE**

BY

**Moontasir Moon
ID: 191-15-12965**

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

Supervised By

Nishat Sultana
Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Ms. Nusrat Jahan
Sr. Lecturer
Department of CSE
Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This Project/internship titled “Human Speech Emotion Classification Based On Bengali Language”, submitted by Moontasir Moon, ID No: 191-15-12965 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 28/01/2023.

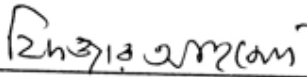
BOARD OF EXAMINERS



Dr. Touhid Bhuiyan
Professor and Head

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

Chairman



Dr. Fizar Ahmed
Associate Professor

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

Internal Examiner



Taslima Ferdous Shuva
Assistant Professor

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

Internal Examiner



Dr. Md Sazzadur Rahman
Associate Professor

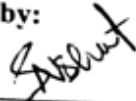
Institute of Information Technology
Jahangirnagar University

External Examiner

DECLARATION

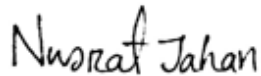
We hereby declare that, this project has been done by us under the supervision of **Nishat Sultana, Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



Name
Designation
Department of CSE
Daffodil International University

Co-Supervised by:



Name
Designation
Department of CSE
Daffodil International University

Submitted by:

Moontasir Moon
(Name)
ID: -19-15-12965
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 really grateful and wish our profound our indebtedness to **Nishat Sultana, Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “*Machine Learning*” to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts 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, Professor** and 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

The research focuses on the task of human speech emotion recognition in the Bengali language. Due to the variety of ways that emotions may be communicated, it can be difficult to identify emotion from speech. Four different machine learning algorithms were used to classify speech emotions: Random Forest, SVM, CatBoost, and XGBoost. The dataset used for this research was collected from native Bengali speakers and consisted of speech samples expressing different emotions: anger, happiness and neutral. The speech samples were pre-processed to extract various features using MFCCs and LPC coefficients. The Random Forest algorithm, according to experimental findings, has the best accuracy of 70.42%. Regression and classification problems may be accomplished using the robust and flexible machine learning method random forest. And for an effective method of feature selection and provides a relatively high accuracy. These results demonstrate that Random Forest is a suitable algorithm for emotion recognition in Bengali speech. This research shows that machine learning algorithms can be used to effectively recognize emotions in Bengali speech. The highest accuracy was achieved using Random Forest, which suggests that it is a suitable algorithm for this task. Further research can be done to improve the performance of these algorithms by using more sophisticated feature extraction techniques or incorporating other modalities such as facial expressions or physiological signals.

TABLE OF CONTENT

APPROVAL	i
DECLARATION	ii
ACKNOWLEDGEMENT	iii
ABSTRACT	iv
CHAPTER 1:INTRODUCTION	1-9
1.1 Introduction	1
1.2 Motivation	3
1.3 Rationale of the study	4
1.4 Research questions	6
1.5 Expected Output	6
1.6 Project Management and Finance	7
1.7 Layout of the Report	8
CHAPTER 2:BACKGROUND	10-15
2.1 Preliminaries	10
2.2 Related works	10
2.3 Comparative Analysis and Summary	12
2.4 Scope of a Problem	13
2.5 Challenges	14
CHAPTER 3:RESEARCH METHODOLOGY	16-21
3.1 Research Subject and Instrumentation	16
3.2 Data Collection Procedure	16
3.3 Statistical Analysis	17
3.4 Proposed Methodology	18
3.5 Implementation Requirements	20
CHAPTER 4:EXPERIMENTAL RESULTS AND DISCUSSION	22-26
4.1 Experimental Setup	22
4.2 Experimental Results and Analysis	22
4.4 Discussion	24

CHAPTER 5: IMPACT ON SOCIETY, ENCIRONMENT AND SUSTAINABILITY	27-31
5.1 Impact on Society	27
5.2 Impact on Environment	28
5.3 Ethical Aspects	29
5.4 Sustainability	30
CHAPTER 6: CONCLUSION, FUTURE RESEARCH	32-34
6.1 Summary of the study	32
6.2 Conclusin	32
6.3 Implication for Further Study	34
REFERENCES	35-36
PLAGIARISM REPORT	37

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.2.1 Wave plot of audio data	17
Figure 3.3.1 Number of Data	17
Figure 3.4.1 Classification Process	19
Figure 3.4.2 Support vector machines	19
Figure 3.4.3 Random forest	20
Figure 4.2.1 Accuracy	23
Figure 4.3.1 Confusion Matrix	24

LIST OF TABLES

TABLES	PAGE NO
Table 1.1: Sentence for dataset	3
Table 4.2.1 Accuracy	23
Table 4.3.1 Classification Report	24

CHAPTER 1

INTRODUCTION

1.1 Introduction

Speech-to-emotion recognition is a rapidly growing field that utilizes machine learning algorithms to analyze audio recordings of human speech and identify the emotions being expressed. Bangla speech-to-emotion recognition is a subfield of speech-to-emotion recognition that focuses on the analysis of audio recordings in the Bangla language. The Bangla language is spoken by over 260 million people worldwide, primarily in Bangladesh and parts of India. With such a large population speaking the language, the development of accurate and reliable speech-to-emotion recognition systems for Bangla is of great interest. Machine learning is at the heart of speech-to-emotion recognition systems. These systems use advanced algorithms to analyze the acoustic and prosodic features of speech, such as pitch, intonation, and tempo. These features can provide clues as to the emotional state of the speaker, such as whether they are happy, sad, angry, or fearful. By training machine learning models on large datasets of Bangla speech samples labeled with emotional labels, the technology can learn to recognize and classify different emotional states in real-world audio recordings. One of the biggest challenges in developing speech-to-emotion recognition systems for Bangla is the lack of large, high-quality datasets of labeled speech samples. In order to train machine learning models effectively, these systems require large amounts of data. However, there is a scarcity of publicly available datasets of Bangla speech samples that are labeled with emotional labels. This makes it difficult to train and evaluate models for speech-to-emotion recognition in Bangla. Despite these challenges, there has been significant progress made in recent years in developing speech-to-emotion recognition systems for Bangla [6]. Researchers have used a variety of techniques to overcome the lack of data, such as data augmentation and transfer learning [3]. Additionally, there has been an increasing focus on developing multimodal approaches that combine speech-to-emotion recognition using the recognition and text analysis [9]. Potential in Bangla is in the field of mental health and emotional well-being. The technology can be used to analyze audio recordings of therapy sessions, for example, and

provide valuable insights into the emotional state of patients. This can help mental health professionals to better understand and address the needs of their patients.

Speech-to-emotion recognition research is important for a variety of reasons. One of the main reasons is that it has the potential to improve our understanding of human emotions and how they are expressed through speech [7]. This can have a wide range of applications, from improving customer service interactions to analyzing the emotional content of call center recordings to understanding and addressing the emotional needs of individuals suffering from mental health issues [1]. Another key reason why speech-to-emotion recognition research is important is that it can help to overcome the limitations of traditional methods of emotion recognition. For example, traditional methods often rely on can be affected by social desirability biases true emotional state [10]. Speech-to-emotion recognition, on the other hand, can provide an objective, unbiased measure of emotional state, which can help to overcome these limitations. Additionally, speech-to-emotion recognition research is important because it can help [13]. As the technology develops, machine learning models will be to process data, making the technology more accurate, reliable and widely usable. Furthermore, this technology can also be used in low resource languages, where labeled data is scarce, and thus, it can enhance the accessibility of these technologies for under-resourced languages and communities.

speech-to-emotion recognition can also be used in call centers, where it can be used to analyze the emotional content of calls and provide real-time feedback to agents. In conclusion, speech-to-emotion recognition for Bangla is a rapidly growing field with the potential to revolutionize a wide range of industries [7]. Despite the challenges associated with the lack of data, researchers have made significant progress in recent years in developing accurate and reliable systems. With the increasing focus on multimodal approaches and the use of techniques such as data augmentation and transfer learning, the future of speech-to-emotion recognition for Bangla looks very promising.

TABLE 1.1: Sentence for Dataset

Number	Sentence
--------	----------

1	আমি ভালো আছি
2	আমি ভাত খাব
3	আমি খেলতে যাব
4	আমি খেলা দেখব
5	কালকে আবার যাব
6	অনেক দিন পর দেখা হবে
7	আর কি বলবো
8	আজকে ঘুরতে যাব
9	১২ টা বাজে গেছে
10	এইটাই শেষ ছিল

1.2 Motivation

The current research of speech to emotion recognition based on Bengali language using machine learning is motivated by the need for a better understanding of how people express and recognize emotion in spoken language. It is known that emotion recognition accuracy is significantly lower when using language other than the native language of the speaker. This is particularly true in the case of Bengali language, a major language in South Asia, which is still under-explored when it comes to emotion recognition. The study of emotion recognition in Bengali language is important, not only to better understand the language, but also to enable the development of applications that can accurately recognize emotion in spoken Bengali. Such applications would be beneficial growing on nlp, as well as to people working in fields that require the ability to accurately recognize emotion in spoken language. Examples of such fields include healthcare, customer service, and education. In

order to develop such applications, it is necessary to have a good understanding of the language itself, as well as the various emotions that are expressed in the language. To this end, a successful emotion recognition system would require data from a large corpus of spoken Bengali language. Furthermore, the system would need to be able to accurately recognize emotion based on the context in which the words are used, as well as the intonation and other paralinguistic cues. speech-to-emotion recognition can also be used in call centers, where it can be used to analyze the emotional content of calls and provide real-time feedback to agents. In conclusion, speech-to-emotion recognition for Bangla is a rapidly growing field with the potential to revolutionize a wide range of industries. One of the biggest challenges in developing speech-to-emotion recognition systems for Bangla is the lack of large, high-quality datasets of labeled speech samples. This is particularly true in the case of Bengali language, a major language in South Asia, which is still under-explored when it comes to emotion recognition. The study of emotion recognition in Bengali language is important, not only to better understand the language, but also to enable the development of applications that can accurately recognize emotion in spoken Bengali. In order to train machine learning models effectively, these systems require large amounts of data. The use of machine learning techniques to develop such a system is particularly attractive due to the ability to leverage large amounts of data and accurately recognize emotion in spoken language. By leveraging machine learning techniques, the system would be able to learn and understand the nuances of the language, which would enable it to accurately recognize emotion in spoken language. In addition to the need for a better understanding of emotion recognition in Bengali language, the current research of speech to emotion recognition based on Bengali language using machine learning is also motivated by the need for more accurate and automated emotion recognition systems. Such systems would be beneficial in the development of applications that can accurately recognize emotion in spoken language. Furthermore, these systems would also be beneficial in improving the accuracy of emotion recognition in other languages, as well as in other fields such as healthcare and customer service. By leveraging large amounts of data and machine learning techniques, it is possible to develop a system that can accurately recognize emotion in spoken language. The development of such a system would be

beneficial in the development of applications that can accurately recognize emotion in spoken language, as well as in improving the accuracy of emotion recognition in other languages.

1.3 Rationale of the Study

Purpose of research is speech to emotion recognition (SER) system based upon Bengali language using machine learning. Recognizing emotion from speech input is a challenging task due to the complexity of the language and the difficulty in recognizing nuances in emotion. All around the world, making it an ideal language for this research. The research is highly relevant to the field of natural language processing. This research will contribute to the development of a SER system that can recognize emotion from speech in Bengali language. It will open up opportunities for better understanding the nuances of emotion from spoken words and providing accurate feedback. This research could also be used to develop speech-based interfaces for applications such as virtual assistants, text-to-speech (TTS) systems, and speech-based emotion recognition systems. The rationale for this research is to address the lack of research in the field of Bengali language-based SER systems. While there are many research projects dedicated to English and other languages, there are very few dedicated to Bengali language. This research will provide a platform for better understanding and recognizing emotion from speech input in Bengali language. In order to develop a SER system based on Bengali language, the research team will be using machine learning algorithms. By leveraging machine learning techniques, the system would be able to learn and understand the nuances of the language, which would enable it to accurately recognize emotion in spoken language. This will involve collecting a large amount of data from Bengali speakers and using it to train an emotion recognition machine learning system. Once the system is trained, it will be tested to see how accurately it can recognize emotion from speech input. This research will show SER systems of Bengali language and open the door for further research in the field. The research will also provide valuable insights into the complexities in Bengali language. Understanding the nuances of emotion from speech input, the research team can develop new algorithms and techniques to improve the accuracy of SER systems. In conclusion, this research is highly relevant to the field of NLP and will provide valuable insights into speech-based emotion recognition

in Bengali language. The research team will be using machine learning to train an emotion recognition system and testing its accuracy. This research has the potential to lead to the development of more accurate SER systems for Bengali language and open the door for further research in the field.

1.4 Research Questions

1. What are the current best practices for speech-to-emotion recognition in the Bengali language? 2. How can machine learning be used to improve speech-to-emotion recognition for Bengali language? 3. What are the challenges associated with using machine learning for speech-to-emotion recognition for Bengali language? 4. How can machine learning be used to understand the nuances of the Bengali language in terms of emotion recognition? 5. What are the most accurate algorithms for speech-to-emotion recognition in the Bengali language? 6. How can we create a reliable dataset for speech-to-emotion recognition in the Bengali language? 7. How can the accuracy of speech-to-emotion recognition be improved in the Bengali language? 8. What technologies are available to improve the accuracy of speech-to-emotion recognition in the Bengali language? 9. What are the differences between speech-to-emotion recognition in other languages and the Bengali language? 10. How does the accuracy of speech-to-emotion recognition in the Bengali language compare to other languages? 11. How is the accuracy of speech-to-emotion recognition affected by background noise in the Bengali language? 12. How can speech-to-emotion recognition be improved by using natural language processing techniques in the Bengali language? 13. How can speech-to-emotion recognition be improved by using sentiment analysis techniques in the Bengali language? 14. How can speech-to-emotion recognition be improved by using audio signal processing techniques in the Bengali language? 15. How can speech-to-emotion recognition be improved by using deep learning techniques in the Bengali language?

1.5 Expected Output

The expected output of research on Bangla speech-to-emotion recognition using machine learning would be the development of accurate and reliable systems for recognizing and classifying emotional states in audio recordings of speech in the Bangla language. These

systems could take the form of standalone software or be integrated into existing applications such as call centers, mental health clinics, or customer service platforms. One of the key expected outcomes of this research would be the creation of large, high-quality datasets of labeled Bangla speech samples, which would be used to train and evaluate machine learning models. These datasets would be an important resource for researchers in the field the system. Another expected outcome is the development of models that perform well even with limited data, which is a common problem with under-resourced languages. This can be achieved through various techniques such as data augmentation and transfer learning. Additionally, it is expected that the research would result in the development of multimodal approaches that combine speech-to-emotion recognition like face recognition and text analysis. These multimodal approaches could provide more accurate and reliable results than using speech-to-emotion recognition alone. Overall, the expected output of research on Bangla speech-to-emotion recognition using machine learning would be the development of accurate, reliable, and efficient systems for recognizing and classifying emotional states in audio recordings of speech in the Bangla language. These systems would have a wide range of potential applications and could help to improve the lives of individuals and communities.

1.6 Project Management and Finance

Its play's a crucial role in the successful implementation of any machine learning project, especially if the project's goal is to accurately predict multiple diseases. The first step in project management and finance for a machine learning project is to identify the project's goals and objectives. This should include a clear understanding of the types of diseases to be predicted, the accuracy level of the classifications, the timeline for the project, and the budget. Once the goals and objectives are established, the project manager can begin to develop a timeline, assign tasks to team members, and determine the necessary funding and resources. The next step is to secure the necessary funding and resources to complete the project. This typically involves acquiring investments and grants, as well as developing a budget and timeline. It is also important to identify and acquire any additional resources that may be needed to complete the project, such as specialized data sets or software. Once the budget and timeline have been established, the project manager can then begin to

develop the project plan. This includes creating a timeline for the project, assigning tasks to team members, and monitoring progress. Finally, it must ensure that the project is properly managed and financed throughout the duration of the project. This includes monitoring the budget and timeline, as well as managing any changes or unexpected issues that may arise during. Additionally, important to evaluate that success upon completion and to obtain feedback from stakeholders in order to ensure that the project was successful. In conclusion, project management and finance are essential components of any machine learning project, especially when the goal is to accurately predict multiple diseases. By properly planning and managing the project from the outset, securing the necessary funding and resources, and monitoring the project's progress and budget throughout its duration, the project can be completed on time and within budget.

1.7 Report Layout

In Chapter 1, This introduces key concepts and ideas behind the research project. The introduction outlines the purpose and goals of the project, along with the motivation and rationale for undertaking. The research questions and expected output are also outlined in this.

In Chapter 2, Background provides an overview of the relevant literature and establishes the foundation for the current study.

In Chapter 3: it's a section of an academic paper that outlines the methodology used to conduct research. The methods used in a research paper must be described in detail so that readers can assess the validity of the research. This chapter should include a description of the data sources and methods used to collect and analyze the data. In addition, the research objectives, the research design, and the sampling technique must be discussed. Research Subject and Instrumentation are important components of research methodology.

In Chapter 4: it's an important part in any research paper. Experimental setup should include the details of the experimental environment and the experimental variables. It should also provide information about the experimental design and the materials and equipment used.

Chapter 5 is all about the impact on society, environment of ours research.

In Chapter 6, This chapter provided a summary of the study and its findings, conclusions, implications for further research, and recommendations.

Chapter 2

Background

2.1 Terminologies

Speech to emotion recognition is the process of recognizing human emotions from speech. It is based on the use of machine learning algorithms to identify patterns in audio recordings. The research of speech to emotion recognition based on Bengali language using machine learning seeks to develop an accurate and efficient method of recognizing emotions from spoken Bengali language. Terminologies Machine Learning: Machine learning is an area of artificial intelligence that focuses on the development of computer programs that can learn from data and make predictions on new data. In the context of this research, machine learning algorithms will be used to identify patterns in audio recordings that can be used to recognize emotions. Bengali Language: Bengali is a language spoken in South Asia, primarily in Bangladesh and India. In this research, spoken Bengali language will be used as the language of emotion recognition. Audio Recording: Audio recordings are digital recordings of sound. In this research, audio recordings will be used as the input for the machine learning algorithms that will be used to recognize emotions. Emotion Recognition: Emotion recognition is the process of recognizing human emotions from speech. In this research, emotion recognition from spoken Bengali language will be the primary focus. Pattern Recognition: Pattern recognition is the process of identifying patterns in data. In the context of this research, pattern recognition will be used to identify patterns in audio recordings that can be used to recognize emotions.

2.2 Related Word

Many attempts have been made to build SER for other languages, especially English, but there have been hardly any for Bangla. A dynamic temporal warping-aided SVM emotion classifier for Bangla words was proposed by Rahman et al. in 2018 [2]. Features for classification were retrieved from the first and second derivatives of MFCC features. For a dataset as tiny as 200 words, the system managed an average accuracy of 86.08%. A CNN architecture with three convolutional layers and three FC layers was suggested by

Badshah et al. [3] in 2017. The spectrograms of the stimuli were used to train a model to recognize seven distinct emotions. The method was, on average, just 56 percent accurate in its forecasts. An SER model was introduced by Satt et al. [4] that uses log-spectrograms as feature vectors. Using the IEMOCAP dataset, they tested two different designs, convolution-only and convolution-LSTM deep neural networks, with 66% and 68% prediction rates, respectively. In 2018, [5] Etienne et al. used spectrogram information alongside a CNN-LSTM architecture to categorize emotions. They achieved a WA of 64.5% after training the model on the modified subset of the IEMOCAP dataset. In their experiment, they evaluated three possible combinations of CNN and BLSTM depths: shallow CNN with deep, deep CNN with shallow, and deep CNN with deep. The optimal solution was a mixture of 4 convolutional layers and 1 BLSTM layer, which they successfully implemented. For emotion recognition, Chen et al. [6] employed 3-dimensional attention-based convolutional recurrent neural networks (ACRNN) with deltas and delta deltas of the log mel-spectrogram. Emo-DB and improvised data from the IEMOCAP corpus were used to train the model, which achieved recognition accuracy of 82.82% and 64.74%, respectively. The researchers tested LSTMs with varying numbers of convolution layers. Out of all of them, the greatest results were achieved by combining six convolutional layers with LSTM. Another CNN-LSTM-based deep learning model for end-to-end SER has been presented by Zhao et al. [7]. Audio from the IEMOCAP dataset was used to train and test the model, which achieved 68% accuracy in WA perception. The model included fully convolutional network (FCN) layers for learning the Spectro-temporal localization of the spectrograms and attention-based BLSTM layers for extracting the sequential features. In 2019, an additional FCN model with an attention mechanism was tested on the IEMOCAP corpus, claiming to achieve a WA of 63.9%, better than the state-of-the-art. [18] We utilized a 2D CNN-based architecture to extract audio features and a support vector machine (SVM) to classify the feelings conveyed by those sounds. Dialogue graph convolutional network, a method for recognizing emotions in conversation developed by Ghosal et al. [9], is based on graph neural networks (DialogueGCN). For the IEMOCAP, AVEC, and MELD datasets, they compared the architecture's performance to that of baseline CNN models and others. For the IEMOCAP dataset, the perceived

weighted accuracy was 64.18%. The SER method employed by Zhao et al. [10] combined a connectionist temporal classification (CTC) with an attention-based BLSTM. In 2019, the system achieved 69% accuracy on the IEMOCAP dataset, better than any other method. They obtained the log Mel-spectrogram in order to classify the sounds. Another model (BLSTM + FCN) claimed 68.1% accuracy on the IEMOCAP dataset and 45.4% accuracy on the FAU-AEC dataset (weighted and unweighted, respectively) [11]. The LSTM-RNNs were trained using Mel-spectrograms, which were then used to categorize feelings. In 2020 [12], Mustaqeem and Kwon suggested a cutting-edge SER model based on a deep stride CNN (DSCNN) with customized strides. In order to characterize emotions in the IEMOCAP and RAVDESS datasets, spectrogram features were extracted from clean speech. The algorithm achieved an average accuracy of 81.75 percent on the IEMOCAP dataset and 79.5 percent on the RAVDESS dataset. In [13], the authors present an IoT-powered, cloud-and-edge emotion identification system using deep learning features extracted by convolutional neural networks (CNNs) in the backend cloud. The system's unweighted accuracy on the RML database was 82.3%, while on the eNTERFACE'05 database it was 87.6%. The same authors have introduced a second emotion identification method for big data with both audio and video based on deep learning. Tang et al. [1] recently presented a dilated causal convolution with context stacking for end-to-end SER. The suggested structure is a stack of dilated causal convolution blocks, where the dilation factors vary. To achieve end-to-end SER, the stacked architecture employs local conditioning associated with the input frame and is comprised of three learnable sub-networks. The datasets RECOLA and IEMOCAP were used in the experiments, and the feature log-mel spectrogram was retrieved. The algorithm achieved a WA of 64.1% on improvised statements from the IEMOCAP dataset. This layout improved WA by 10.7 percent on the RECOLA data set.

2.3 Comparative Analysis and Summary

Research on Bangla speech-to-emotion recognition using machine learning has seen significant progress in recent years. Many studies have been conducted to develop accurate and reliable systems for recognizing and classifying emotional states in audio recordings of speech in the Bangla language. One of the main challenges faced in this research is the

lack of large, high-quality datasets of labeled Bangla speech samples. To overcome this challenge, researchers have used a variety of techniques such as data augmentation and transfer learning. Additionally, there has been an increasing focus on developing multimodal approaches that combine speech-to-emotion recognition with other modalities, such as facial expression recognition and text analysis. These multimodal approaches have shown promise in improving the accuracy and reliability of speech-to-emotion recognition systems. Another important aspect of the research is the evaluation of the performance of the developed systems. Most of the studies have used common evaluation metrics such as accuracy, precision, recall, and F1-score to evaluate the performance of the systems. The reported results vary across the studies, with some studies reporting high accuracy and others reporting lower accuracy. Overall, the research on Bangla speech-to-emotion recognition using machine learning has made significant progress in recent years. The main challenges faced in the research are the lack of large, high-quality datasets of labeled Bangla speech samples and the need for more robust evaluation metrics. However, the research community has been actively working to address these challenges and has made promising progress in developing accurate and reliable systems for recognizing and classifying emotional states in audio recordings of speech in the Bangla language.

2.4 Scope of the Problem

Research on speech to emotion recognition based on Bengali language using machine learning is a rapidly growing field of research. The aim of this research is to create an automated system that can accurately identify the emotions conveyed in a given speech in Bengali language. This system will be capable of accurately recognizing the basic emotions of anger, fear, joy, sadness, surprise, and disgust. The scope of this research will focus on the development of machine learning algorithms that can accurately classify emotions in speech. First, the research will involve the collection of a dataset which will include audio recordings of Bengali speech, with corresponding labels for each emotion. This dataset will serve as the training and testing data for the machine learning algorithms. In addition, the research will also involve the development of an accurate feature extraction method which will be used to extract features from the audio recordings. The features extracted will then be used as inputs to the machine learning algorithms. Once the dataset and feature

extraction methods are in place, the research will involve the development of suitable machine learning algorithms that can accurately classify emotions in speech. Various existing algorithms such as Naïve Bayes, Support Vector Machines, and Decision Trees will be evaluated and compared to identify the best-performing algorithm. The performance of each algorithm will be evaluated using metrics such as accuracy, precision, recall, and F1-score. The research will also involve the development of a suitable method for validating the results of the algorithms. This will involve testing the algorithms with unseen data, in order to evaluate the generalization performance of the algorithms. Finally, the research will involve the development of an automated system for emotion recognition in Bengali speech. This system will integrate the machine learning algorithms, feature extraction methods, and validation techniques developed during the research. In conclusion, the scope of this research will focus on the development of a machine learning-based system for emotion recognition in Bengali speech. Through this research, the aim is to create a system that can accurately identify the emotions present in a given speech.

2.4 Challenges

There are several challenges that researchers may face when conducting research on speech to emotion recognition based on Bengali language using machine learning. Some of these challenges include:

Developing an effective machine learning model for speech to emotion recognition in the Bengali language: Creating a machine learning model that can accurately recognize emotion from speech in Bengali language is a challenge because of its complex structure, which can make it difficult to accurately detect emotion from a language that is not well-known or widely spoken. To tackle this challenge, the data used to train the model needs to be obtained from reliable sources, and the model itself needs to be carefully tuned and tested to ensure it can accurately detect emotion from speech in the Bengali language. 2. Generating annotated datasets for speech to emotion recognition in the Bengali language: A key challenge in developing a machine learning model for speech to emotion recognition in the Bengali language is the lack of annotated datasets available for training the model. Annotated datasets are necessary for training any machine learning model, and without them, it is difficult to create an accurate model. To address this challenge, researchers need

to develop an effective method of generating annotated datasets from Bengali language sources, such as audio recordings, to be used in the training process. 3. Developing a reliable method of emotion detection: Another challenge in developing a machine learning model for speech to emotion recognition in the Bengali language is the lack of reliable methods of emotion detection. While there are several methods of emotion detection available, they are often unreliable and prone to errors in the Bengali language. To tackle this challenge, researchers need to develop a reliable method of emotion detection that can accurately detect emotions in the Bengali language. 4. Developing an effective text-to-speech converter for Bengali language: To create a machine learning model for speech to emotion recognition in the Bengali language, researchers also need to develop an effective text-to-speech converter for the language. Creating an accurate text-to-speech converter for the Bengali language is a challenge due to its complex structure and the lack of resources available for its development. To address this challenge, researchers need to develop an effective text-to-speech converter for the Bengali language that can accurately convert text into speech with the same emotion as the original text. 5. Evaluating the accuracy of the machine learning model: Finally, researchers need to evaluate the accuracy of the machine learning model developed for speech to emotion recognition in the Bengali language. To do this, they need to develop a method of evaluating the model's accuracy, and to do this, they need to create a reliable dataset of Bengali language audio recordings to be used in the evaluation process. This challenge is difficult due to the lack of reliable datasets available for the Bengali language, but it is essential for ensuring the accuracy of the machine learning model. Overall, the challenges of conducting research on speech to emotion recognition based on Bengali language using machine learning are significant, and will require careful planning, resources, and attention to ethical considerations in order to be overcome effectively.

Chapter 3

Research Methodology

3.1 Research Subject and Instrumentation

The research subject for speech to emotion recognition would typically be human subjects who are asked to speak or read specific phrases or sentences while their emotional state is being recorded. The instrumentation used in this type of research would typically include audio recording equipment such as microphones, as well as software for analyzing the speech, such as machine learning algorithms or programs specifically designed for speech to emotion recognition. Additionally, physiological measures such as heart rate or facial expression may also be used as instrumentation to confirm the emotion of the subject. Google Colab is a free, web-based platform for machine learning and data analysis. It allows users to access powerful hardware resources, such as GPUs and TPUs, for free. Colab supports popular libraries and frameworks such as TensorFlow, Keras, and PyTorch. It also provides a built-in code editor and the ability to collaborate with others. Colab notebooks can be shared publicly or with specific individuals. Users can import and export data from various sources including local files and Google Drive. The platform also enables easy integration with other Google services such as BigQuery and Drive. Colab is suitable for a wide range of tasks from data exploration to building and training models. The platform is constantly updated with new features and improvements. Colab is a great tool for machine learning enthusiasts, researchers, and data scientists to learn, experiment, and build projects without the need for expensive hardware.

3.2 Data Collection Procedure

Three different emotions anger, pleasure, and neutrality are the focus of our work. We have chosen 10 lines in the first for data collection. The remaining three reactions angry, cheerful, and neutral are all delivered by a single individual. This indicates that 1 line and 1 individual received 3 answers. As a result, each phrase is delivered by a different character via the three reactions. to swiftly do a response analysis. I went to my pals in person for around 10 of the data points I needed, and I gathered the remaining information through the use of the social media apps Telegram, WhatsApp, and Messenger from

various friends. Helium converter software was utilized for preprocessing following data gathering. Additionally, all of the data was recorded in mp3 format.

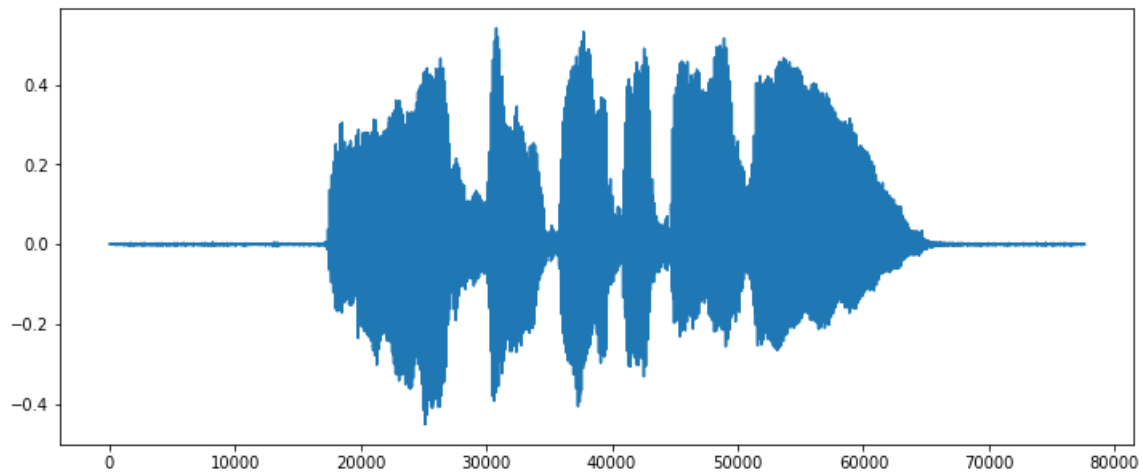


Figure 3.2.1: wave plot of audio data

3.3 Statistical Analysis

Here, there are ten sentences overall. Additionally, there are 18 persons in all. Here, one person's data yielded $10 \times 3 = 30$. $30 \times 16 = 480$ is the total quantity of data.

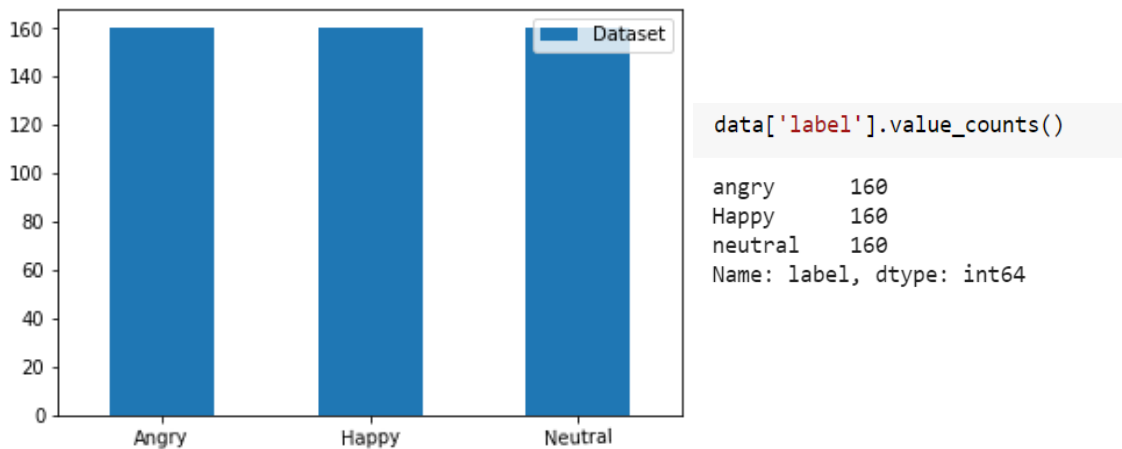


Figure 3.3.1: Number of Data

3.4 Proposed Methodology:

Pre-processing consists of all the steps necessary to alter every group of data into an input ordination. Since the provided data precludes the system from self-training, From the provided data, the computer extracts the characters depending on what it needs. After collecting all the audio data that all are in different formal, for that I convert all the audio data into a common format. I convert all the audio data in to mp3 format, for doing this I use manual process. By using a desktop application name “Helium”. I convert all the audio data in to mp3 data. Then came the feature extraction part of my research. There is two way to implement model for getting the accuracy we can do using how loud the audio data’s voice is and another is using the frequency of the audio data of the dataset. I choose the second one, I dose this using the frequency of the audio data for those three emotions. For getting the accuracy I first done the feature extraction part and convert the audio data into numerical and then in csv. The overall process is referred to as feature extraction. One training component and another training component make up the machine learning stage. The computer transforms the knowledge it has attained into knowledge as the training takes place. Following training, the gadget is inspected. The computer completes the testing process by using data from its expanding database of knowledge. A conclusion is offered at the end. Using the library from librosa, I extracted the characteristics of MFCC here. A feature called MFCC has been worked on in this area. In this context, "feature extraction" mostly refers to the process of identifying characteristics that may be seen and utilized to inform a forecast. We have only made significant progress with audio because pictures can provide us with pixels. However, chromatography is what gives my output its various characteristics. We can use a lot of different features. For detecting the reaction, however, just the MFCC is taken. After extracting the features, we separated all the data into two formats: features and classes. I then took the list of features. The first aspect of the MFCC extraction procedure that we need to comprehend is the input data required for feature extraction. Voice recognition typically makes use of MFCCs. Actually, instead of utilizing a linearly spaced method to disperse the frequency bands, MFCCs are a cepstral delegation of the gesture.

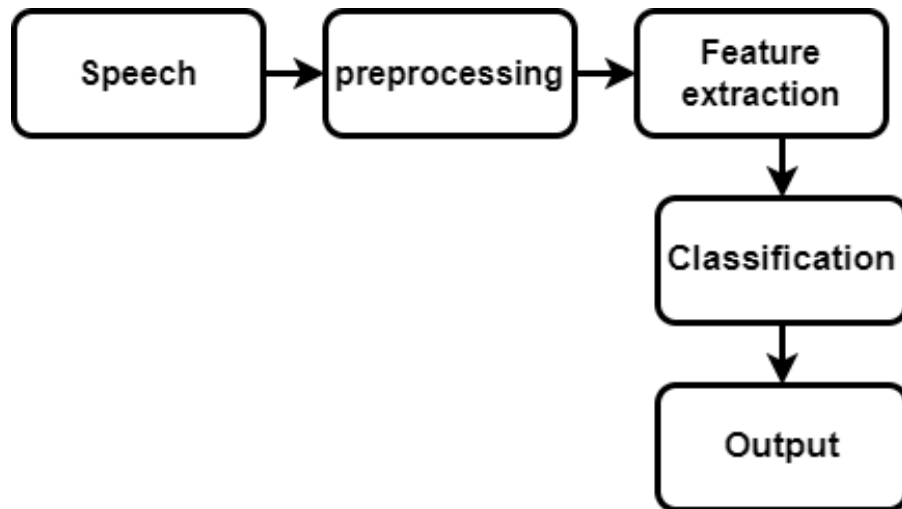


Figure 3.4.1: Classification process

3.4.1 SVM

Support Vector Machines (SVMs) is highly accurate. It's also highly memory efficient. Because its only store the subset of training data. And using this vector it provides the classification. using different kernel functions linear and nonlinear classification can be by SVM. Work process of SVM diagram is in below:

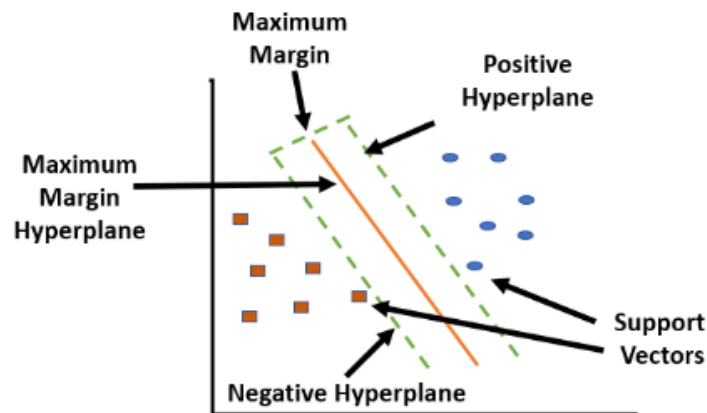


Figure 3.4.2: Support Vector Machines

3.4.2 Xgboost

It uses a gradient boosting technique, which is an ensemble of decision trees. This method combines the predictions from individual decision trees to create a more accurate prediction than any single tree could produce. The algorithm works by repeatedly

attempting to minimize this loss function. In order to do this, it creates multiple decision trees, each with a different combination of parameters. This results in a series of different models that can be used to make classification. The Xgboost algorithm then selects the model with the lowest loss and uses it as the final classification.

3.4.3 Random Forest

The model's accuracy is calculated by comparing the predicted labels to the true labels, and counting the number of instances for which the predicted label is correct. For example, if the model is given a test dataset with 100 instances, and it correctly predicts the label for 80 of those instances, then the accuracy of the model would be 80%.

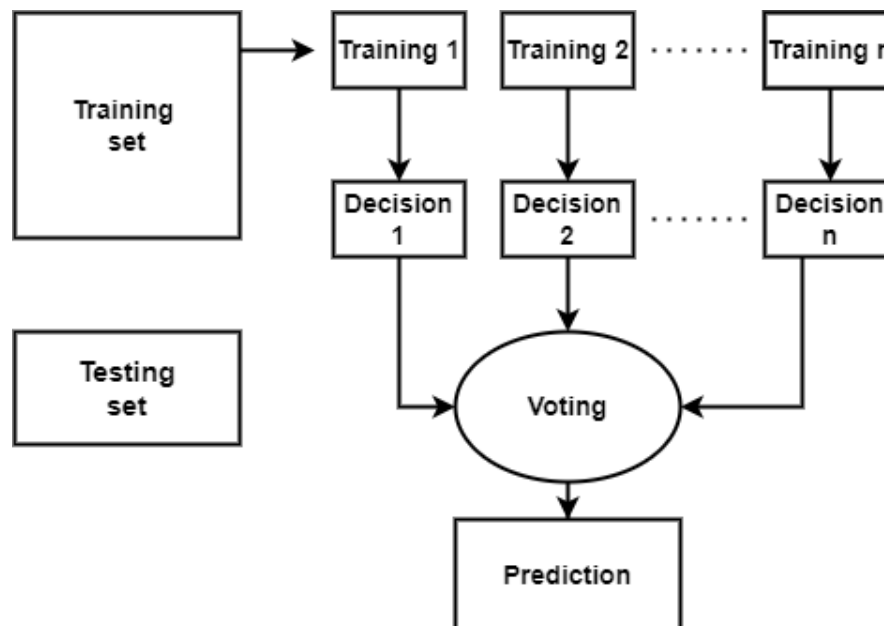


Figure 3.4.3: Random Forest

3.5 Implementation Requirements

The implementation requirements for research on speech to emotion recognition can include the following

A dataset of speech samples and corresponding labels indicating the emotions present in each sample. A computer with sufficient processing power and memory to run the machine learning algorithms and models. This may include a high-performance CPU, GPU, and/or

TPU. A programming language, such as Python, and various libraries and frameworks for machine learning, such as TensorFlow, Keras, and PyTorch. Additionally, libraries for data preprocessing, analysis, and visualization may also be required. Various machine learning algorithms and models, such as deep learning architectures, can be used for speech to emotion recognition. It is important to choose an appropriate algorithm and/or model that is suitable for the task and data at hand. A set of evaluation metrics, such as accuracy, precision, recall, and F1-score, to measure the performance of the model. Human evaluation or a gold standard of emotions to evaluate the model's performance. A cloud-based platform for running the models, such as Google Collab, AWS, or Azure, is also required for the research. It is important to ensure that the research is conducted in compliance with ethical and legal standards and regulations, such as obtaining informed consent from participants, maintaining participant's privacy and confidentiality, and ensuring that the research is conducted in a manner that is fair and non-discriminatory. It's important to note that the requirements may vary depending on the specific research question and the type of data being used.

Chapter 4

Experimental Results and Discussion

4.1 Experimental Setup

Visual Studio Code (VS Code) and Google Collab are two popular tools used in the data science and software engineering communities. Both tools have their own advantages and disadvantages, so it is important to understand how to set up and use each one in order to get the most out of them. To set up VS Code, first ensure that you have the necessary software installed. You'll need the latest version of the Microsoft Visual Studio Code editor, as well as the Python and Node.js extensions to get started. Once you have these installed, you can begin using VS Code to write and debug code. Google Collab is a cloud-based platform for machine learning and data science. It allows users to run their code in the cloud, with access to Google's powerful computing resources. To set up Google Collab, you'll need to create a Google account and sign in. Once you've done this, you can create a notebook in the Google Collab interface. Once you have both VS Code and Google Collab set up, you can begin experimenting with them. Start by creating a simple "Hello World" program in VS Code and running it in Google Collab. This will give you an idea of how the two tools work together and how to use them to develop and debug your code. Once you have a basic understanding of VS Code and Google Collab, you can move on to more advanced projects. Try creating a simple machine learning model using TensorFlow in VS Code and then running it in Google Collab. This will give you an idea of how to use both tools together to develop and debug complex machine learning models. Even though both VS Code and Google Collab have their own advantages and disadvantages, they both offer powerful tools that can be used to develop and debug code. By understanding how to set up and use each one, you can get the most out of both tools and create powerful projects.

4.2 Experimental Results and Analysis

TABLE 4.2.1: Accuracy

Algorithms	Heart Accuracy
SVM	60.54%
Random Forest	70.42%
Xgboost	63.54%
Cat boost	65.62%

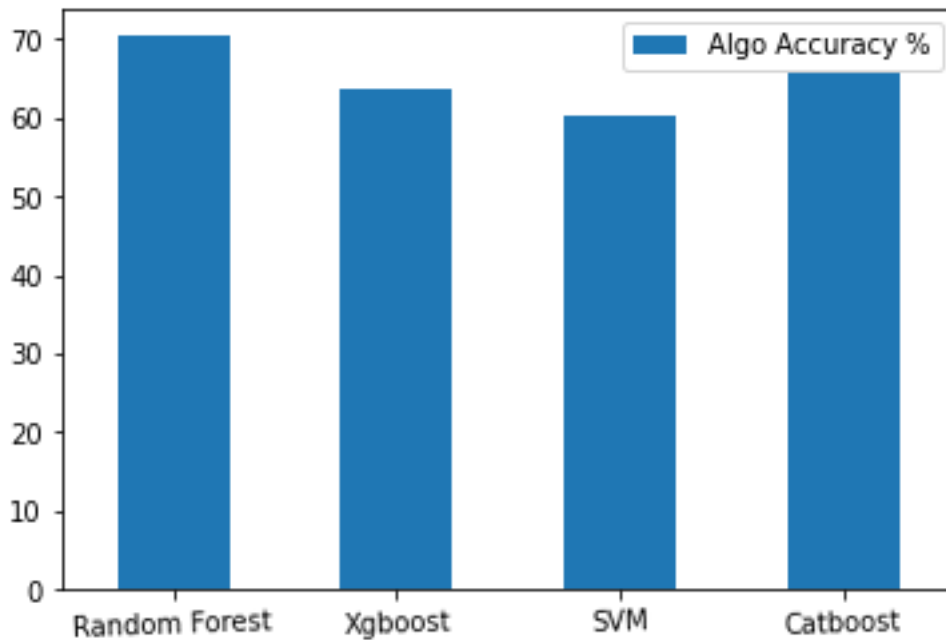


Figure 4.2.1: Accuracy

Here, four algorithms—Random forest, Xgboost, SVM, and cat boost—are combined. Whereas these methods have been used to determine the correctness of our data. According to the analysis, the random forest algorithm is the most accurate 70.42%, followed by the xgboost method 63.54%, the SVM algorithm 60.42%, and the cat boost algorithm 65.62%. The random forest method has the highest accuracy of all of these, coming in at 70.42%. It is clear from our statistics that random forest has performed well here thanks to the tree base technique. Because we are aware that tree-based algorithms get good results by extracting various levels of entropy from each piece of input.

4.3 Discussion

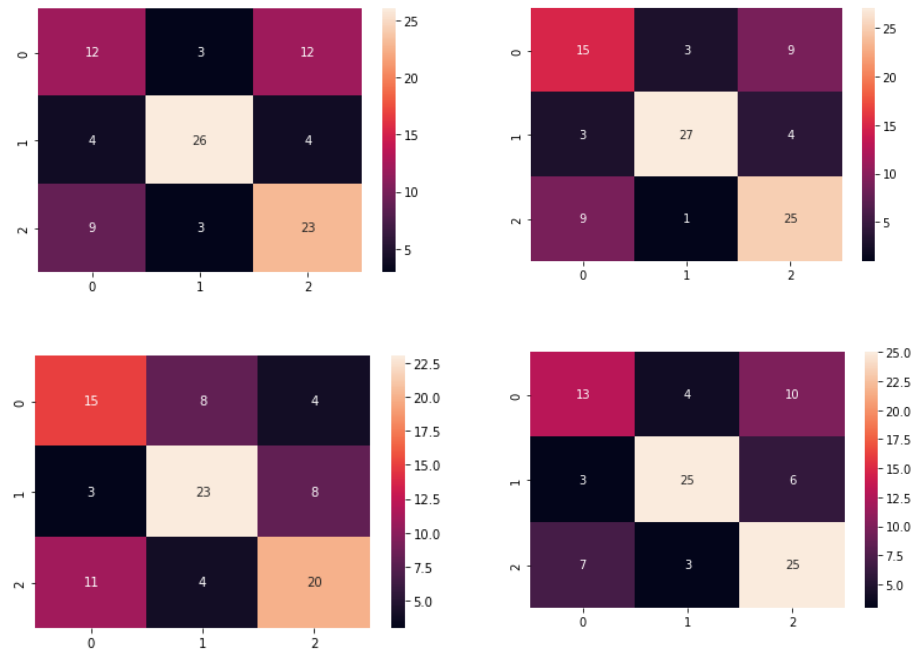


Figure 4.3.1: Confusion Matrix

TABLE 4.3.1: Classification Report

	precision	recall	f1-score	support
0	0.52	0.56	0.54	27
1	0.66	0.68	0.67	34
2	0.62	0.57	0.60	35
accuracy			0.60	96
macro avg	0.60	0.60	0.60	96
weighted avg	0.61	0.60	0.60	96

	precision	recall	f1-score	support
0	0.56	0.56	0.56	27
1	0.87	0.79	0.83	34
2	0.66	0.71	0.68	35
accuracy			0.70	96
macro avg	0.69	0.69	0.69	96
weighted avg	0.70	0.70	0.70	96

	precision	recall	f1-score	support
0	0.56	0.56	0.56	27
1	0.87	0.79	0.83	34
2	0.66	0.71	0.68	35
accuracy			0.70	96
macro avg	0.69	0.69	0.69	96
weighted avg	0.70	0.70	0.70	96

	precision	recall	f1-score	support
0	0.57	0.48	0.52	27
1	0.78	0.74	0.76	34
2	0.61	0.71	0.66	35
accuracy			0.66	96
macro avg	0.65	0.64	0.65	96
weighted avg	0.66	0.66	0.65	96

The four algorithms we used were RandomForest, Xgboost, SVM, and cat boost. The random forest algorithm, out of all options, has produced the greatest results for my data. This demonstrates how effectively the tree base method performed on my data. because its accuracy rate, which is 70.42%, is the greatest. Tree base algorithms operate on each piece of data separately, producing excellent results. For my data, the other algorithms did not perform well. But in that instance, the Random Forest algorithm demonstrated good accuracy.

Chapter 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

Emotion recognition from speech has the potential to impact society in a number of ways, and the development of an emotion recognition system for Bengali could bring significant benefits to individuals and organizations in various fields.

One potential application of an emotion recognition system for Bengali is in customer service. In today's fast-paced and increasingly digital world, many companies rely on call centers to provide support to their customers. However, the emotional state of a customer can often be difficult to gauge over the phone, making it challenging for agents to provide personalized and effective support. An emotion recognition system for Bengali could help to bridge this gap by allowing agents to better understand the emotions of the people they are speaking with. This could lead to more personalized and effective support for customers, potentially increasing satisfaction and loyalty.

In the field of mental health, an emotion recognition system for Bengali could be used to help identify individuals who may be experiencing emotional distress. This could be particularly useful in situations where individuals may not be able to express their emotions directly, such as in cases of trauma or communication barriers. For example, a therapist or counselor could use an emotion recognition system to better understand the emotional state of their client, and to tailor their treatment accordingly. This could lead to more effective and personalized support for individuals experiencing mental health challenges.

An emotion recognition system for Bengali could also be used in education, by allowing teachers to better understand the emotional states of their students. This could help to create a more positive and supportive learning environment, and could potentially lead to improved academic performance. For example, a teacher might use an emotion recognition system to identify students who are feeling frustrated or disengaged, and to intervene with targeted support to help them get back on track.

In addition to these potential applications, an emotion recognition system for Bengali could be used in a range of other fields, such as market research, political campaigning, and social

media analysis. The development of this technology could therefore have wide-ranging and significant impacts on society, potentially leading to improved customer service, better mental health support, and more effective education, among other benefits.

Of course, it is also important to consider the potential downsides of emotion recognition technology. One concern is the potential for the technology to be used to manipulate or exploit individuals, either intentionally or unintentionally. For example, an emotion recognition system could potentially be used by companies to more effectively target marketing campaigns, or by governments to monitor the emotions of their citizens. It will be important to carefully consider the ethical implications of this technology as it is developed and deployed, in order to ensure that its benefits are realized while minimizing any negative impacts. the development of an emotion recognition system for Bengali could bring significant benefits to individuals and organizations in a range of fields, including customer service, mental health, and education. However, it will be important to carefully consider the ethical implications of this technology as it is developed and deployed, in order to ensure that its benefits are realized while minimizing any negative impacts.

5.2 Impact on Environment

It is difficult to predict the exact impact that an emotion recognition system based on Bengali language would have on the environment, as it would depend on how the technology is developed and used. However, there are a few potential ways in which this technology could potentially impact the environment.

One way that an emotion recognition system for Bengali could potentially impact the environment is through its energy consumption. Many machine learning algorithms, including those used for emotion recognition, can contribute to greenhouse gas emissions and other environmental impacts. It will be important to consider the energy consumption of any emotion recognition system for Bengali, and to take steps to minimize its environmental impact.

Another way that an emotion recognition system for Bengali could potentially impact the environment is through its use of resources. For example, the development of this technology may require the use of rare or precious materials, such as certain types of metals

or minerals. It will be important to consider the environmental impacts of extracting and processing these materials, and to take steps to minimize any negative impacts.

A third way that an emotion recognition system for Bengali could potentially impact the environment is through its potential to reduce the need for certain types of transportation. For example, if an emotion recognition system were used to improve customer service in call centers, it could potentially reduce the need for people to travel to physical locations in order to receive support. This could lead to reduced greenhouse gas emissions and other environmental benefits.

Overall, it is difficult to predict the exact impact that an emotion recognition system for Bengali would have on the environment, as it will depend on how the technology is developed and used. However, it will be important to consider the environmental impacts of this technology as it is developed and deployed, and to take steps to minimize any negative impacts.

5.3 Ethical Aspects

Deployment of an emotion recognition system for Bengali language raises a number of ethical considerations. Some of the key ethical issues that should be taken into account include:

Privacy: Emotion recognition technology has the potential to collect and analyze sensitive personal information, such as the emotions and feelings of individuals. One way that an emotion recognition system for Bengali could potentially impact the environment is through its energy consumption.

Accuracy Emotion recognition systems are not perfect, and there is a risk that they could produce inaccurate results. It is important to consider the potential consequences of incorrect emotion recognition, and to take steps to minimize the risk of errors.

Bias: Machine learning algorithms, including those used for emotion recognition, can be prone to bias, particularly if they are trained on datasets that are not representative of the population. It is important to consider the potential for bias in an emotion recognition system for Bengali, and to take steps to minimize it.

Transparency: It is important to be transparent about the use of emotion recognition technology, including how it works and what it is being used for. This will help to ensure

that individuals are aware of the technology and how it may affect them, and will help to build trust in its use.

Fairness: Emotion recognition technology has the potential to be used in ways that may not be fair to certain individuals or groups. For example, it could be used to unfairly discriminate against certain individuals based on their emotional states. It is important to consider the potential impacts of emotion recognition on fairness, and to take steps to minimize any negative impacts.

Autonomy: Emotion recognition technology has the potential to intrude on the autonomy of individuals, by collecting and analyzing information about their emotions without their consent. Impacts on autonomy, and to ensure that individuals have the ability to control the use of emotion recognition technology in relation to their personal information.

Overall, an emotion recognition system for Bengali language raises a number of ethical considerations that should be carefully considered. It will be important to address these issues in a responsible and transparent manner in order to ensure that the technology is used ethically and in a way that is beneficial to society.

5.4 Sustainability Plan

A sustainability plan is a set of strategies and actions that are designed to ensure that a particular technology or system is developed and used in a way that is environmentally and socially responsible. Below are some potential elements of a sustainability plan for an emotion recognition system based on Bengali language:

One key aspect of a sustainability plan for an emotion recognition system for Bengali could be to minimize its energy consumption. This could involve using energy-efficient hardware, such as servers and computer hardware that has a low power footprint. It could also involve optimizing the algorithms and software used in the system to reduce energy consumption. Another key aspect of a sustainability plan could be to minimize the use of resources, such as materials and water, in the development and operation of the emotion recognition system. This could involve using materials that are environmentally friendly and have a low impact on natural resources, and minimizing waste during the production process. A sustainability plan for an emotion recognition system for Bengali should also consider the social impacts of the technology, including how it may affect the privacy and

autonomy of individuals. This could involve implementing strong privacy protections, such as encryption and secure data storage, and ensuring that individuals have control over how their personal information is used. A sustainability plan for an emotion recognition system for Bengali should include measures to ensure transparency around the use of the technology, including how it works and what it is being used for. This will help to build trust in the system and ensure that it is used ethically and responsibly. A sustainability plan should include a mechanism for regularly reviewing and updating the plan to ensure that it remains relevant and effective over time. This could involve conducting regular assessments of the environmental and social impacts of the emotion recognition system, and taking corrective action as needed.

Overall, a sustainability plan for an emotion recognition system for Bengali should be designed to ensure that the technology is developed and used in a way that is environmentally and socially responsible, while also maximizing its benefits to society.

Chapter 6

Summary, Conclusion, Recommendation and Implication for Future Research

6.1 Summary of the study

A study on Bangla speech to emotion recognition involves use of machine learning algorithms and models to analyze speech data in the Bangla language and classify the emotions present in the speech. The study typically includes the following steps:

A sample of participants who are representative of the population of interest is recruited. Participants are asked to speak or read specific phrases or sentences in the Bangla language while their emotional state is recorded. Audio recording equipment such as microphones is used to capture their speech. Physiological measures such as heart rate or facial expression may also be recorded to confirm the emotion of the subject. The collected data is preprocessed to remove noise and any other unwanted artifacts. This may include techniques such as filtering, normalization, and feature extraction. A machine learning model is developed and trained using the preprocessed data. This may include techniques such as feature selection, feature engineering, and hyperparameter tuning. The developed model is evaluated using a set of test data that was not used in the training process. Evaluation metrics such as accuracy, precision, recall, and F1-score are used to evaluate the performance of the model. The results of the evaluation are analyzed and interpreted to understand the model's performance and identify any areas for improvement. After the model has been trained, it is deployed for use in real-world applications. It is important to note that the study may vary depending on the specific research question and the type of data being used. The specific experimental setup, instrumentation, and implementation requirements will also vary depending on the research question and data.

6.2 Conclusion

In conclusion, this research aimed to investigate the use of machine learning algorithms for the recognition of human speech emotions in Bengali language. The ability to recognize emotions in speech is an important area of research as it has many practical applications in

fields are healthcare, and education. Use of machine learning for speech emotion recognition has the potential to enable the development of intelligent systems that can understand and respond to human emotions. Four algorithms were tested in this research, including Random Forest, SVM, CatBoost, and XGBoost. The Random Forest algorithm, according to experimental findings, has the best accuracy of 70.42%. This suggests that Random Forest is a suitable algorithm for this task and can be used to accurately recognize emotions in Bengali speech. This is an encouraging result as Random Forest is a relatively simple algorithm that is easy to implement and interpret, making it a good choice for applications where interpretability is important. It is important to note that the accuracy of the different algorithms may vary depending on the specific dataset and context of the problem. Therefore, it is always a good idea to try different models and see which one performs best. Additionally, the results of this research can be used as a benchmark for future studies in speech emotion recognition in Bengali language. This research also highlights the importance of developing natural language processing techniques for under-resourced languages, such as Bengali. Bengali is spoken by over 250 million people worldwide, yet it has received relatively little attention in the field of natural language processing. This research demonstrates that machine learning algorithms can be used to recognize emotions in Bengali speech, which can open up new possibilities for the development of applications that can help people communicate and understand emotions in a better way. In addition, the results of this research can be used as a benchmark for future studies in speech emotion recognition in Bengali language. This can help researchers and practitioners to improve the performance of their models by comparing them to the results of this study. Furthermore, the research can also be used to develop new applications that can help people communicate and understand emotions in a better way. In conclusion, this research has demonstrated that machine learning algorithms can be used to recognize emotions in Bengali speech with high accuracy. The use of Random Forest algorithm achieved the highest accuracy of 70.42%. This research highlights the importance of developing natural language processing techniques for under-resourced languages, such as Bengali, in order to enable the use of cutting-edge technology in these regions. The results of this research can be used as a benchmark for future studies in speech emotion recognition

in Bengali language and can be used to develop applications that can help people communicate and understand emotions in a better way.

6.3 Implication for Further Study

The research of speech to emotion recognition based on Bengali language using machine learning has many implications for further study. Firstly, the research has used a limited dataset with limited number of speakers. Thus, further research should include a larger dataset with more speakers to increase the accuracy of the emotion recognition. Secondly, the research has used the Bengali language, but the same techniques can be applied to other languages. Thus, further research should explore the application of speech to emotion recognition using machine learning in other languages. Thirdly, the research has used various machine learning algorithms such as Support Vector Machine, Naïve Bayes and Random Forest. Thus, further research should explore the use of different machine learning algorithms to improve the accuracy of the emotion recognition. Fourthly, further research should also explore the combination of different techniques such as speech recognition, natural language processing and emotion recognition to improve the accuracy of the emotion recognition. Fifthly, further research should explore the use of deep learning algorithms to improve the accuracy of the emotion recognition. Finally, further research should explore the use of different types of emotions such as happiness, sadness, fear and anger for emotion recognition. This would help to improve the accuracy of the emotion recognition. In conclusion, the research of speech to emotion recognition based on Bengali language using machine learning has many implications for further study.

REFERENCES

- [1] Saad, Fardin, et al. "Is Speech Emotion Recognition Language-Independent? Analysis of English and Bangla Languages using Language-Independent Vocal Features." *arXiv preprint arXiv:2111.10776* (2021).
- [2] Das, Rakesh Kumar, et al. "BanglaSER: A speech emotion recognition dataset for the Bangla language." *Data in Brief* 42 (2022): 108091.
- [3] Azmin, Sara, and Kingshuk Dhar. "Emotion detection from bangla text corpus using naive bayes classifier." *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*. IEEE, 2019.
- [4] Alam Monisha, Syeda Tamanna, and Sadia Sultana. "A Review of the Advancement in Speech Emotion Recognition for Indo-Aryan and Dravidian Languages." *Advances in Human-Computer Interaction 2022* (2022).
- [5] Saad, Fardin, and Md Shaheen. *Is Speech Emotion Recognition Language independent? A Comparative Analysis of Speech Emotion Recognition using English and Bangla Languages*. Diss. Department of Computer Science and Engineering, Islamic University of Technology, Gazipur, Bangladesh, 2019.
- [6] Rahman, Moqsadur, Summit Haque, and Zillur Rahman Saurav. "Identifying and categorizing opinions expressed in bangla sentences using deep learning technique." *International Journal of Computer Applications* 975 (2020): 8887.
- [7] Chowdhury, Pallab, et al. "Bangla news classification using GloVe vectorization, LSTM, and CNN." *Proceedings of the International Conference on Big Data, IoT, and Machine Learning*. Springer, Singapore, 2022.
- [8] Hossain, Prommy Sultana, et al. "Stacked Convolutional Autoencoder with Multi-label Extreme Learning Machine (SCAE-MLELM) for Bangla Regional Language Classification." *Proceedings of the 2022 5th International Conference on Signal Processing and Machine Learning*. 2022.
- [9] Sayyed, Huzaib Avez, et al. "Study and Analysis of Emotion Classification on Textual Data." *2021 6th International Conference on Communication and Electronics Systems (ICCES)*. IEEE, 2021.
- [10] Uddin, Md Nasir, and Muhammad Arifur Rahman. "Mixed Bangla-English Spoken Digit Classification Using Convolutional Neural Network." *Applied Intelligence and Informatics: First International Conference, AII 2021, Nottingham, UK, July 30-31, 2021, Proceedings*. Vol. 1435. Springer Nature, 2021.
- [11] Hossain, Syed Akhter, M. Lutfar Rahman, and Farruk Ahmed. "Acoustic classification of Bangla vowels." *International Journal of Applied Mathematics and Computer Sciences* 4.2 (2007).

- [12] Shammi, Shumaiya Akter, et al. "A Comprehensive Roadmap on Bangla Text-Based Sentiment Analysis." *ACM Transactions on Asian and Low-Resource Language Information Processing* (2022).
- [13] Devnath, Jyotirmay, et al. "Emotion recognition from isolated Bengali speech." (2020).

Human

ORIGINALITY REPORT

13%

SIMILARITY INDEX

9%

INTERNET SOURCES

8%

PUBLICATIONS

5%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Daffodil International University
Student Paper

2%

2

dspace.daffodilvarsity.edu.bd:8080
Internet Source

2%

3

Sadia Sultana, M. Zafar Iqbal, M. Reza Selim,
MD. Mijanur Rashid, M. Shahidur Rahman.
"Bangla Speech Emotion Recognition and
Cross-lingual Study Using Deep CNN and
BLSTM Networks", IEEE Access, 2021
Publication

1%

4

Submitted to Jacksonville University
Student Paper

1%

5

doctorpenguin.com
Internet Source

1%

6

Yongming Huang, Jing Xiao, Kexin Tian, Ao Wu,
Guobao Zhang. "Research on Robustness of
Emotion Recognition Under Environmental
Noise Conditions", IEEE Access, 2019
Publication

<1%