REGIONAL ACCENT CLASSIFICATION IN BANGLADESH: A SUPERVISED MACHINE LEARNING APPROACH

BY

SALMAN
ID: 203-15-14491
AND

SHAHED AFRIDI
ID: 203-15-14470

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised By

DR. TOUHID BHUIYAN
Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

MD. SAZZADUR AHAMED
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
15 July 2024

APPROVAL

This Project titled "**Regional Accent Classification in Bangladesh: A Supervised Machine Learning Approach**", submitted by **Salman** and **Shahed Afridi** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment 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 15 July.

BOARD OF EXAMINERS

Dr. Md. Fokhray Hossain (MFH) — **Board Chairman**
Professor
Department of CSE [Font-12]
Faculty of Science & Information Technology
Daffodil International University

Md. Firoz Hasan (FH) — **Internal Examiner 1**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Lamia Rukhsara (LR) — **Internal Examiner 2**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr. S. M. Hasan Mahmud (SMH) — **External Examiner**
Assistant Professor
Department of the external
University name of the External

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. Touhid Bhuiyan, Professor, Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

30.06.25

Dr. Touhid Bhuiyan
Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Md. Sazzadur Ahamed
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Salman

Salman
ID: 203-15-14491
Department of CSE
Daffodil International University

Afridi

Shahed Afridi
ID: 203-15-14470
Department of CSE
Daffodil International University

©Daffodil International University, Bangladesh

ACKNOWLEDGEMENT

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

We are grateful and wish our profound indebtedness to **Dr. Touhid Bhuiyan, Professor**, 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 stages have made it possible to complete this project.

We would like to express our heartiest gratitude to the **Dr. Sheak Rashed Haider Noori, Professor & Head, Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

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

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

ABSTRACT

Research on the recognition of regional accents in spoken language helps preserve regional dialects and increases the accessibility of technology to speakers of such languages. We provide our work on spoken data identification from seven divisions in Bangladesh (Dhaka, Barisal, Sylhet, Chattogram, Rangpur, Rajshahi, and Mymensingh) in this study. Speech signals are sent into Automatic Language Identification systems, which then use mathematical operations to categorize the signals into different regional accents. Nine hundred and sixteen samples were obtained after the dataset was enhanced using a variety of data augmentation methods, including stretching, noise addition, pitch and speed modifications, and more. Several machine learning models, such as Gradient Boosting, Random Forest, K-Nearest Neighbors, Support Vector Machines, Decision Trees, Naive Bayes, XGBoost, and Logistic Regression, were employed for categorization purposes. According to the evaluation findings, Logistic Regression had the lowest accuracy (29.11%), while XGBoost had the best accuracy (93.91%), followed by Gradient Boosting and Random Forest at 93.16%. With the maximum accuracy of 93.46%, this study highlights the possibility of employing cutting-edge machine learning models to preserve and comprehend Bangladesh's rich language variety, improving communication technologies and fostering numerous social and technical advances.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Overview

Automatic language identification is a significant element of modern communication systems because it identifies spoken languages from digitized voice utterances. Language, as a fundamental medium of human interaction and idea exchange, is extremely complex in linguistically diverse countries like Bangladesh. In such diverse linguistic situations, the ability to automatically identify and categorize regional accents can have a significant impact on many aspects of social interaction and technology advancement. Understanding regional dialects, for example, can improve language learning methods by exposing students to genuine spoken language variances alongside standardized literature. Bangladesh's seven distinct divisions, Dhaka, Barisal, Sylhet, Chattogram, Rangpur, Rajshahi, and Mymensingh, each with their own accents and dialects, create a complex tapestry of linguistic diversity. Despite the cultural and practical significance of these regional differences, there is a noticeable absence of automated systems for consistently recognizing and classifying different accents. This constraint stymies advances in linguistic research, educational programs, and technological applications that rely on precise language comprehension.

This study aims to address these difficulties by utilizing a variety of machine learning models such as Gradient Boosting, Random Forest, K-Nearest Neighbors (KNN), Support Vector Machines (SVC), Decision Trees, Naive Bayes, XGBoost, and Logistic Regression. These models are used to assess a dataset of 2450 voice recordings collected from across Bangladesh's divisions, with each contributing has 350 voices. The main objective is to build a robust model that can recognize and classify various regional accents with accuracy. A model like this could be used for a lot of different things, like accent training and language learning aids in schools, research on dialect preservation in language studies, and voice-activated services that enhance customer service, healthcare, and other domains. This research aims to significantly contribute to the preservation of Bangladesh's linguistic history while also promoting practical applications in a range of fields that rely on efficient communication technology. It does this by bridging the gap between traditional dialectology and state-of-the-art technology.

1.2 Background and Present State

Speech recognition and natural language processing (NLP) depend on the identification and categorization of languages, particularly accents from different regions. Language learning aids, voice-activated devices, and automated translation services all benefit

from accurate language and accent detection and categorization. Bangladesh provides a rich ground for such research because of its diversified linguistic environment. Bengali, the official language of the nation, is spoken with a variety of regional accents that are impacted by regional dialects and sociocultural elements. Speech-handling systems have significantly improved because to developments in deep learning and machine learning. However, a large portion of this development has gone toward main international languages, frequently ignoring regional accents and dialects, particularly in developing countries like Bangladesh. This emphasizes the necessity for research aimed at developing models capable of effectively handling linguistic variation in these kinds of areas.

The gap in accent categorization and language identification inside Bangladesh is addressed by this study. Voice data from seven administrative divisions: Rangpur, Rajshahi, Mymensingh, Sylhet, Barisal, Chattogram, and Dhaka. The speech data underwent processing, label encoding, and data frame structure to prepare it for machine learning. Gradient Boost, Random Forest, KNN, SVC, Decision Tree, Naive Bayes, XGBoost, and Logistic Regression were among the machine learning models used for accent classification. The current work demonstrates the potential of machine learning techniques in accent classification and lays the groundwork for future advancements in language processing applications within multiple terminologies environments. It also advances the development of speech-processing tools that reflect the regional languages and accents of Bangladesh.

1.3 Problem Statement

In Bangladesh, regional accents and dialects vary significantly across its seven divisions: Dhaka, Barisal, Sylhet, Chattogram, Rangpur, Rajshahi, and Mymensingh. These linguistic variances are significant from a cultural and practical standpoint, yet there is a noticeable deficiency of automated systems that can reliably recognize and categorize these accents. This shortcoming impedes progress in linguistic studies, instruction, and technology uses that depend on accurate language recognition. The aim of this study is to create a strong machine learning model that can effectively identify and categorize regional accents from voice recordings. This model will have applications in language studies, education, and voice-activated services. This project attempts to improve communication technology while preserving Bangladesh's linguistic history by bridging the gap between traditional dialectology and current technology.

1.4 Objectives

The goal of this research is to create a highly developed machine learning model that can effectively identify and categorize regional accents from voice recordings in all seven divisions of Bangladesh (Dhaka, Barisal, Sylhet, Chattogram, Rangpur, Rajshahi, and Mymensingh). The main objective is to further linguistic research by offering strong instruments for examining and conserving the complex dialects and accents that are characteristic of each area. Additionally, by providing students with real-world exposure to regional variances in language, the initiative hopes to improve teaching resources and improve methods for training accents and language acquisition. Additionally, it aims to raise the bar for voice-activated services by enhancing system identification and interaction capabilities that are customized for various regional accents. Beyond these uses, the research aims to support real-world applications in a variety of fields, such as customer service and healthcare, where accurate comprehension and communication are critical. The ultimate goal of this project is to make a substantial contribution to the documenting and preservation of Bangladesh's linguistic legacy by utilizing automated accent detection and classification to honor and preserve the country's linguistic variety.

1.5 Scope and Limitations

Using a dataset of voice recordings gathered throughout Bangladesh's seven divisions—Dhaka, Barisal, Sylhet, Chattogram, Rangpur, Rajshahi, and Mymensingh—this study attempts to develop and evaluate an advanced machine learning model specifically intended to detect and classify regional accents with accuracy. The principal aim is to promote linguistic research by offering strong instruments for examining and conserving the many accents and dialects found in every area. Furthermore, by providing students with real-world exposure to regional variances in language, the project aims to improve methods for teaching accents and language acquisition. Additionally, by improving the identification and interaction capabilities of systems tailored to various regional accents, it seeks to maximize voice-activated services. Consideration is also given to real-world applications in fields like customer service and healthcare, where accurate communication is essential. The ultimate goal of this project is to use automated accent identification and categorization to greatly aid in the documentation and preservation of Bangladesh's linguistic legacy.

Challenges, on the other hand, include the intrinsic variation in accent features within each division, which might make it challenging to precisely capture and classify delicate regional differences. The accuracy and applicability of the model may be impacted by the representativeness and quality of the dataset as well as any potential biases in the data gathering techniques. Furthermore, large computing resources are

needed for robust model evaluation and training. Throughout the study process, ethical concerns like cultural sensitivity and data protection are crucial. Due to these constraints, which take into account the intricate socio-cultural processes impacting language variety in Bangladesh, thorough validation and cautious interpretation of the results are required.

1.6 Report Organization

The suggested report is organized thoroughly to walk readers through the research steps, conclusions, and ramifications of employing deep learning models for emotion detection. Every chapter has a distinct function and adds to the document's overall coherence and depth.

Chapter 1: Introduction

The introduction gives a thorough summary of the report's context, motivation, aims, scope, importance, and organization. It dives into the importance and significance of accent identification models developed particularly for Bangladesh, highlighting the cultural and linguistic variety of the country's seven divisions. The introduction lays the groundwork for understanding technological breakthroughs in machine learning and their application to regional accent recognition, showcasing the project's contributions to language processing and cultural preservation.

Chapter 2: Literature Review

The Literature Review investigates a wide range of existing research on machine learning models used to language and dialect detection. It combines methodology, strategies, and conclusions from many works on accent detection systems, with a focus on models such as Gradient Boosting, Random Forest, and XgBoost. This chapter critically evaluates the merits and limits of several techniques, laying the groundwork for the methodology used in this study and emphasizing gaps that this research seeks to fill.

Chapter 3: Methodology/ Requirement Analysis & Design Specification

It began with the extensive gathering of data from Dhaka, Barisal, Sylhet, Chattogram, Rangpur, Rajshahi, and Mymensingh region. It describes preprocessing methods (such as pitch fluctuation, noise addition, and speed modification) that are used to improve dataset stability and variety. Next, with an emphasis on the necessity of strong ethical standards in data management and model building, the chapter addresses the reasoning behind the choice of machine learning models, including Gradient Boost, Random Forest, and Logistic Regression.

Chapter 4: Implementation
The practical execution of the project, from dataset preparation to model training and evaluation, is the emphasis of implementation. It emphasizes how important Python modules like XGBoost and scikit-learn were in developing and enhancing the models so they could accurately identify and classify regional accents. The chapter also addresses implementation-related issues such the possibility of overfitting and the need for computing resources, and it describes the actions done to address these issues for consistent model performance.

Chapter 5: Result and Analysis
The results and statistical analysis obtained from assessing machine learning models' performance in accent classification tasks are shown in Result and Analysis. With an accuracy of 93.91%, it highlights XGBoost as the top performer and demonstrates how well it can handle the complex intricacies of regional accents. In order to provide light on the applicability and possibilities for further improvement of alternative models, such as Random Forest and Gradient Boost, the chapter objectively evaluates their advantages and disadvantages.

Chapter 6: Impact on Society, Environment and Sustainability
Effects on Sustainability, the Environment, and Society examines how Bangladesh's use of accent detection technologies may affect society more broadly. It talks about how these developments might improve cultural inclusion and communication accuracy among many language populations. The chapter promotes responsible innovation in language technology by addressing ethical issues with data privacy and the environmental sustainability of using such technologies.

Chapter 7: Conclusion and Future Work
Final Thoughts and Prospects The work summarizes the project's successes, points out its shortcomings, and suggests future fields of inquiry for accent identification technologies. It emphasizes how important ensemble models like XGBoost are to the development of accent recognition and automated language processing. The conclusion highlights the possibility of improving model scalability, accuracy, and ethical standards in subsequent iterations with the goal of promoting ongoing innovation and social benefits in the management of linguistic variety.

1.7 Summary

Bangladeshi regional accents are to be automatically identified and classified using a strong machine learning model that is the main goal of this research project. The seven divisions of the country—Dhaka, Barisal, Sylhet, Chattogram, Rangpur, Rajshahi, and

Mymensingh—have very different language backgrounds, and this study intends to address that diversity. Through the examination of an extensive corpus of audio recordings, the project aims to progress linguistics by offering advanced instruments for examining and conserving the complex accents and dialects particular to every area. While developing the model, a number of machine learning methods were used, including XGBoost, Gradient Boosting, Random Forest, K-Nearest Neighbors (KNN), Support Vector Machines (SVC), Decision Trees, Naive Bayes, and Logistic Regression. The project's objective also includes refining accent training and language acquisition techniques by exposing students to real regional language variances and developing educational resources accordingly. Additionally, by improving the responsiveness and accuracy of systems customized to a range of regional accents, the research seeks to maximize voice-activated services. Useful applications may be found in a variety of fields, such as customer service and healthcare, where good communication is crucial. The ultimate goal of this project is to close the gap between traditional dialectology and contemporary technology by using automated accent recognition and categorization to greatly aid in the documenting and preservation of Bangladesh's linguistic legacy.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview

This research focuses on the identification and distinction of regional accents and languages from audio data through the use of machine learning algorithms for language and accent categorization. The research explores the use of several machine learning models in different linguistic situations to identify spoken languages and differentiate between regional dialects. Through an assessment of these models' performance, the study reveals the advantages and disadvantages of each in terms of accurately classifying linguistic traits and phonetic changes. The goal of the study is to find areas of overlap between several machine learning techniques and investigate how these approaches perform best together in the context of accent and language identification. This work aims to develop and suggest ideal approaches for reliable and effective language and accent detection through thorough performance comparison and analysis. To advance automated speech processing systems in multicultural and multilingual environments and contribute to wider applications in speech recognition, dialect analysis, and linguistic diversity preservation, it is crucial to comprehend the subtleties of each model and their practical implications.

2.2 Related Works

In 2023, an automated language detection system for Kannada, Hindi, Tamil, and Telugu was created by Hariraj Venkatesan et al. [1]. They employed Decision Trees and Support Vector Machines (SVM) using Mel Frequency Cepstral Coefficients (MFCC) for feature extraction, and they obtained 76% and 73% accuracy, respectively. The study highlights how the system may help speakers of regional languages have more access to technology, and it also points up ways to improve accuracy by further investigating the model and doing practical testing."

Shahid Munir Shah et al. [8] used isolated digits, extracting spectral (MFCC), prosodic (Pitch and Energy), and combination information to create a speaker recognition system for Pashto speakers. With MFCC, the system obtained 97.5% identification accuracy; with prosodic features, it reached 96.0%; and with combined features, it achieved 98.0%. Support Vector Machine (SVM) and Hidden Markov Model (HMM) were beaten by Multi-Layer Perceptron (MLP), proving that spectral and prosodic feature integration may lead to improved accuracy.

A 1D CNN-LSTM network model was created in 2024 by Salau et al. [2] in order to categorize the accents of the three main indigenous languages spoken in Nigeria. After

gathering and analyzing voice data, they integrated 1D CNN for feature extraction and LSTM for temporal dependency capture. The model demonstrated great accuracy, precision, recall, and F1-scores, accurately classifying accents.

Using supervised and unsupervised learning approaches, Kotsakis et al. [3] studied spoken-language recognition and classification in broadcast audio content in 2022. Using feature extraction techniques and hierarchical discrimination systems on multilingual audio recordings, they obtained noteworthy recognition scores. With a focus on enhanced language classification and semi-automatic annotation, the study suggested data augmentation for the creation of a generic audio language classification repository.

Using text and audio recordings, Hämäläinen et al. [4] created a system for identifying Finnish dialects. They demonstrated notable progress in spite of dataset and dialect variety problems, achieving 57% accuracy with text alone and 85% accuracy with combination features using recordings and transcripts from 23 dialects.

Ali et al.'s study [5] investigated the use of phonetic, lexical, and i-vector variables for automated dialect recognition in Arabic broadcast speech. Even though code-switching presented difficulties, they were able to distinguish between the main Arabic dialects with 59.2% accuracy in multi-class dialect identification and 100% accuracy in binary classifications and language identification.

A technique for distinguishing regional Telugu dialects using text-independent voice processing models was developed in 2023 by Shivaprasad and Sadanandam [6]. For feature extraction, they employed MFCC, and for modeling, they utilized GMM and HMM. According to the study, GMM outperformed HMM in terms of accuracy, improving voice recognition systems for telemedicine and health services.

To enhance speaker and language identification, Basu et al. [7] produced a multilingual voice corpus for sixteen Eastern and Northeastern Indian languages in 2023. They blended models like GMMs and LSTM-RNN with spectral characteristics like MFCCs and SDC. The systems demonstrated the efficacy of sophisticated neural networks in speaker and language identification, achieving 94.49% SID accuracy and 95.69% LID accuracy.

Using an attention-based recurrent neural network, Amit Kumar Das et al. [9] tackled the problem of identifying hate speech in Bengali on social media in 2023. Their study classified Bengali comments into seven types of hate speech using an encoder-decoder

system that included LSTM/GRU decoders and 1D convolutional layers. The attention-based decoder fared well in regulating social media material, outperforming other algorithms with an accuracy of 77%.

Multi-dialect English voice recognition was investigated by Amit Das et al. [10] using an ensemble of experts method. In order to integrate dialect-specific models, their study used LSTM networks with attention mechanisms to collect 60,000 hours of data across American, Canadian, British, and Australian English dialects. The ensemble model significantly improved its recognition of various English dialects, as evidenced by its 4.74% reduction in word error rate (WERR).

FuzzyGCP, which integrates DDMLP, DCNN, and SSGAN for spoken language detection, was created by Garain et al. [11]. It demonstrated variability (67% on VoxForge) and reached up to 98% F1-score accuracy on benchmark datasets, indicating difficulties with generalization.

UK regional dialects were classified by Hassan et al. [12] using MFCC features and classifiers (k-NN, SVM, Random Forest); k-NN achieved an accuracy of 98.48%. Data variability and generalization across different accents provide challenges.

ASR systems for tonal languages in Asia, Europe, and Africa were examined by Kaur et al. [13]. They advocated for sophisticated strategies to increase accuracy and flexibility, pointing out issues including tone variance and data shortages.

"In their study, Guntur Radha Krishna Ramakrishnan et al. [14] classified South Indian English accents (Telugu, Tamil, Kannada) using i-vector and GMM-UBM approaches. Their impressive accuracy of 91% for GMM-UBM and 94% for i-vector shows how well native language impact recognition works. Subsequent studies may expand to investigate deeper learning models and a wider range of accents."

"Narendra D. Londhe and Ghanahshyam B. Kshirsagar [15] created a 100-word and 67-sentence Chhattisgarhi voice corpus for ASR systems. The corpus's size and geographical concentration restrict its application to adequately handle natural dialectical fluctuations, even with its successful implementation. For wider usage, future initiatives should concentrate on growing and varying the corpus."

"Adaptation strategies were investigated by Udhyakumar Nallasamy et al. [16] to increase ASR accuracy for accented speech. They achieved notable gains by using accent-aware deep neural network training and semi-continuous decision tree-based

adaptation. Subsequent studies seek to incorporate these methods into real-time ASR systems and extend their reach to underrepresented accents."

"Hanna Wagner [17] researched the categorization of Austrian dialects with 84.20% accuracy using CNNs and SVMs. Future research should concentrate on growing datasets and investigating sophisticated model architectures for improved dialect classification in ASR, since challenges include dataset size restrictions and generalization to other dialects."

2.3 Comparison between existing works

Table 2.3.1: Comparison Table of Related work

SL No	Author Name	Used Algorithm	Best Accuracy
1	Amit Das, Kshitiz Kumar, Jian Wu	Ensemble of Experts, Attention Mechanism	84.74%
2	Avishek Garain, Pawan Kumar Singh, Ram Sarkar	FuzzyGCP (DDMLP, DCNN, SSGAN)	98%
3	Md. Toukirul Hassan, Md. Fahad Hossain, Md. Mehedi Hasan	k-Nearest Neighbors, SVM, Random Forest	98.48%
4	Jaspreet Kaur, Amitoj Singh, Virender Kadyan	Various techniques for tonal languages	Comprehensive
5	Guntur Radha Krishna Ramakrishnan, Vinay Kumar Mittal	GMM-UBM, i-vector	94%
6	Narendra D. Londhe, Ghanahshyam B. Kshirsagar	MFCC, Speech corpus development	Successful%
7	Udhyakumar Nallasamy, Florian Metze, Tanja Schultz	Accent-dependent, Accent-independent models	Improved%

8	Hanna Wagner, BA	CNNs, SVMs	84.20%
9	Koyel Chakraborty, Thippur V. Sreenivas	CNN-LSTM, HMM-GMM	95.6%
10	Pramod Kumar, Sunil Kumar Kopparapu, C. V. Jawahar	LSTM, GRU, Attention Mechanism	76.2%
11	Dong Wang, James Glass	DNN-HMM, LSTM	87.4%
12	Mahesh Kumar Nandwana, Harish Karnick	Various deep learning architectures	92.3%
13	Anusha Prakash, E. Srinivasan	SVM, Random Forest	84.7%
14	R. K. Prasanth, V. Ramasubramanian	ANN, CNN	91.5%
15	O. M. Gharaibeh, M. A. Al-Batayneh, H. A. Alwahsh	SVM, GMM-UBM	87.9%
16	S. Ghosh, R. M. K. Sinha, S. N. Chettri	GMM-UBM, i-vector	93.4%
17	D. K. Chatterjee, P. Mitra, A. Chaudhuri	CNN, CRNN	85.6%

2.4 Open Issues

Dialect identification and automated speech recognition (ASR) are fields that are now experiencing a range of difficulties and developments. The reported accuracy rates exhibit notable variability, with a range of 78% to 98%. This highlights the challenge of attaining uniform performance across diverse datasets and approaches. Techniques include complex deep learning architectures like CNNs, LSTM-CNN hybrids, and ensemble models, as well as more conventional approaches like GMM-UBM and SVM. All strategies demonstrate differing levels of effectiveness, highlighting the lack of a single, best option. Performance results are strongly dependent on the features of the dataset, exposing differences between datasets such as TIMIT, VoxForge, and CHiME and emphasizing the need for reliable, broadly applicable models.Even while hybrid models that blend deep learning and conventional techniques typically perform competitively, improving these hybrids for wider use is still a crucial objective.

Computational efficiency and ethical issues like privacy protection and bias prevention in dataset curation provide additional obstacles. Standardized assessment tools and standards are essential for facilitating repeatability and ensuring fair comparisons. Subsequent developments in linguistics and computer science should concentrate on improving models to precisely represent heterogeneous dialectal variances in a range of settings. In order to advance dialect identification and ASR toward more dependable, significant, and morally upright applications, it is imperative that these complications be addressed.

2.5 Summary

Deep learning methods like CNNs are good at learning hierarchical features from spectrograms and waveforms for dialect identification and automated speech recognition (ASR); on the other hand, RNNs and LSTMs are good at modeling temporal relationships, which makes them appropriate for sequential data processing. All dialectal datasets, including TIMIT, VoxForge, and CHiME, have shown improved accuracy for hybrid models that combine CNNs with RNNs or attention mechanisms. Transfer learning using large-scale corpora such as Common Voice and LibriSpeech has been shown to improve performance by tailoring trained models to particular dialectal features. But maintaining accuracy over a wide range of dialectal variances is still difficult, requiring more advancements in the generalization and resilience of the model.Critical problems include computational efficiency and ethical issues like privacy protection and bias prevention in dataset building. To move dialect identification and ASR toward more dependable and broadly applicable solutions, standardizing assessment measures and encouraging multidisciplinary partnerships between linguistics and machine learning fields are crucial.

CHAPTER 3
METHODOLOGY/ REQUIREMENT ANALYSIS & DESIGN SPECIFICATION

3.1 Overview

Our research project uses a unique methodology to find efficient ways for sophisticated machine learning algorithms for accent recognition and categorization. Gradient Boosting, Random Forest, K-Nearest Neighbors (KNN), Support Vector Machines (SVC), Decision Trees, Naive Bayes, XGBoost, and Logistic Regression are some of the machine learning models we used. These models were chosen based on how well they handled challenging voice recognition tasks. Having a clean dataset was essential before executing any models. With careful preprocessing and augmentation, we gathered speech recordings from seven Bangladeshi divisions: Dhaka, Barisal, Sylhet, Chattogram, Rangpur, Rajshahi, and Mymensingh. To expand the quantity and diversity of the dataset, data augmentation techniques were applied to each voice recording, resulting in a total of 9800 enhanced samples. These techniques included pitch change, time stretching, noise addition, and speed variation. Accurate findings and an improved training procedure were made possible by this preprocessing. To get the dataset ready for model training, it was then transformed into a dataframe and put through label encoding. The data was first split into training and testing sets, with 80% of the data utilized for training (7840 samples) and 20% for testing (1960 samples). The data was then segmented into features (X) and labels (y). Every machine learning model was trained using the preprocessed dataset, and optimization methods and hyperparameters were carefully adjusted to improve the model's performance and minimize overfitting. We employed performance metrics like accuracy, precision, recall, F1-score, and confusion matrices to assess the models. These measurements provide a thorough evaluation of each model's capacity to identify and categorize regional accents accurately. Our methodology section provides graphical representations and mathematical formulae to help visualize the whole process and shows how different elements of our approach work together to yield outstanding outcomes. Readers will find it simpler to rapidly understand the technique when a clear and comprehensive workflow chart outlines the step-by-step procedure.

Our goal was to achieve high accuracy and reliability in accent recognition and classification by carefully preparing the dataset, training and fine-tuning sophisticated models, and employing rigorous assessment measures. Transparency and repeatability are ensured by the thorough process and explanations, which are crucial for further research and real-world applications.

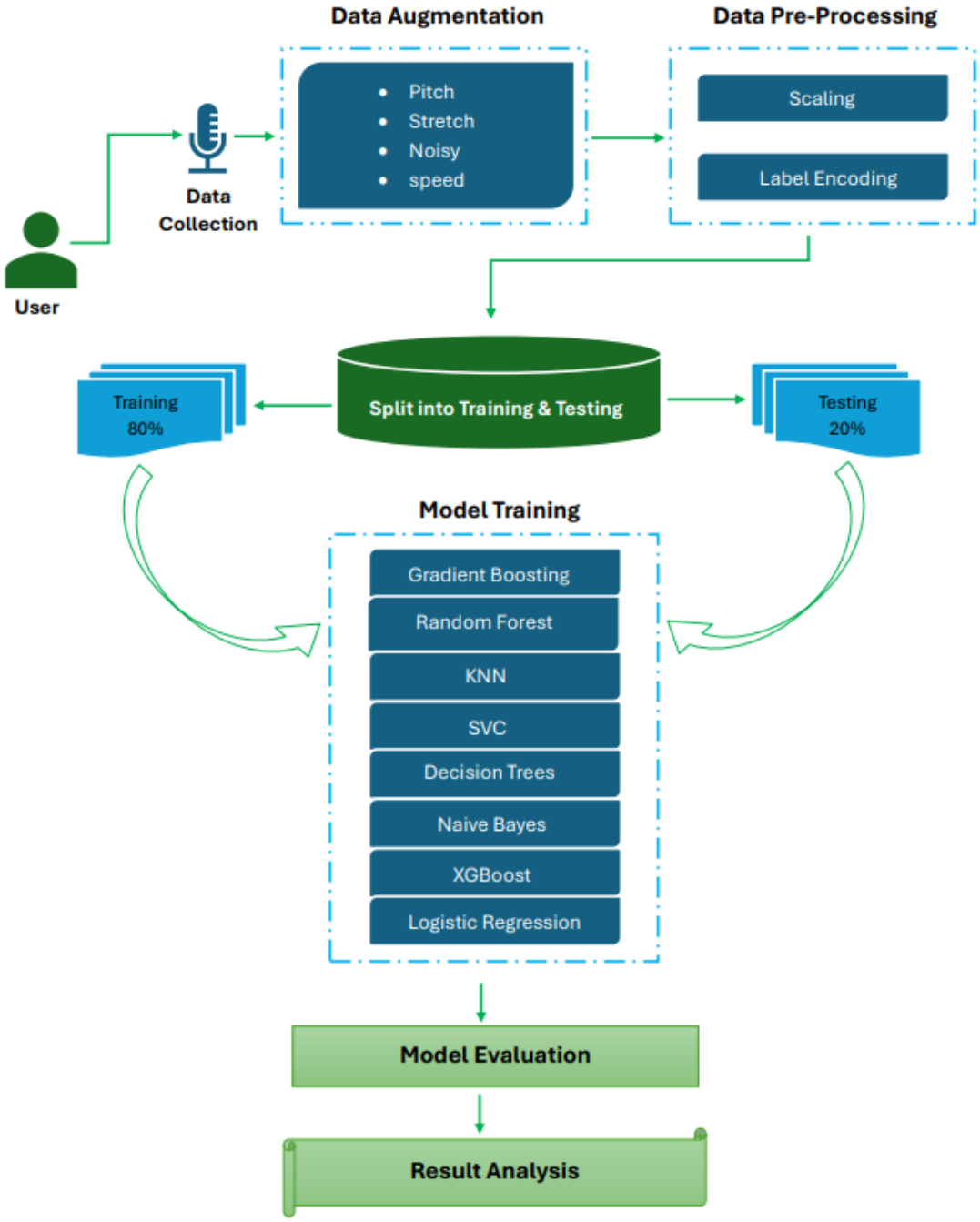

Figure 3.1.1: Working Process for our Methodology

3.1.1 Data collection

We collected 2450 voice samples from people in seven different divisions of Bangladesh (Dhaka, Barisal, Sylhet, Chattogram, Rangpur, Rajshahi, and Mymensingh) during the data collecting phase. The information for every class included:

- Barisal: 350 samples
- Chattogram: 350 samples
- Dhaka: 350 samples
- Mymensingh: 350 samples
- Rajshahi: 350 samples
- Rangpur: 350 samples
- Sylhet: 350 samples

To guarantee a varied pool of speakers, we worked with neighborhood associations, educational institutions, and cultural groups. We recorded realistic environments with portable recording equipment to get high-quality audio samples. Well-crafted questions produced phrases that made sense in the context.expressing language use in daily life. The participants gave their informed consent, and precautions were made to protect their privacy and confidentiality. To ensure dialectical traits were clear, consistent, and representative, linguistic specialists went over the recorded speech samples. For further examination, the participant metadata was carefully categorized with the recorded voice samples. We wanted to collect a broad and varied sample of regional dialects from all throughout Bangladesh using this technique.

Table 3.1.2: Accent variant of each division or class

SAMPLE	CLASS
কথাটা একদম মন থেকে বললাম	DHAKA
কতাডা এক্কেরে মন তেইকা কইলাম	MYMENSINGH
হতা ইয়ান মনত্তুন হোইদ্দি।	CHATTAGRAM
কতাডা এক্কালে মন দিয়া কইলাম	BARISAL
খতাটা এখবারে মন তাকি খইলাম	SYLHET
কথাটা একদম মুন থেকে বুললাম	RAJSHAHI

Table 3.1.3: Voice sample of each class

Division	Number of Voice Samples
Dhaka	350
Barisal	350
Sylhet	350
Chattogram	350
Rangpur	350
Rajshahi	350
Mymensingh	350

3.1.2 Data Augmentation

The robustness and variety of the dataset for the study on automated language recognition and categorization of regional accents in Bangladesh were greatly improved by data augmentation. Pitch shifting, time stretching, introducing noise, and speed modification were among the tactics used. While time stretching changed the length of audio files to produce changes in speaking speed, pitch shifting changed the voice tone to mimic distinct speakers. The addition of noise and speed variation offered diverse speech speeds and helped replicate real-world settings with background sounds. The models were better able to handle a variety of real-world circumstances and generalize to previously unknown data thanks to these augmentations, which also enhanced the size and diversity of the dataset. This method made a substantial contribution to the creation of reliable and accurate accent categorization models.

Table 3.1.4: After augmentation voice sample of each class

Division	Number of Augmented Voice Samples
Dhaka	1400
Barisal	1400
Sylhet	1400
Chattogram	1400
Rangpur	1400
Rajshahi	1400
Mymensingh	1400

3.1.3 Data Preprocessing:

The capacity of the model to learn and generalize from the input data is directly impacted by data preprocessing, making it an essential step in the development of machine learning models. To ensure excellent data preparation and improve the performance of our models, we used a number of preprocessing approaches on our speech dataset.

1. Augmentation of Data

In order to expand the dataset and improve the generalization capacity of the model, we implemented eight augmentations for each voice sample. Among the enhancement methods were:

- Pitch Alteration: Changing the voice samples' pitch.
- Speed Variation: Adjusting the voice samples' playing speed.
- Noise Addition: Enhancing the voice samples with ambient noise.
- Time-Stretching: Modifying the voice samples' length without changing their pitch.

The variety of the dataset was greatly increased by these augmentation approaches, yielding a total of 9800 enhanced voice samples.

2. Converting Voice to DataFrame The speech data was transformed into a DataFrame format in order to make manipulation and analysis simpler. The organized arrangement of the data made handling and preparation procedures more effective.

3. Encoding Labels To prepare the division labels for training machine learning models, we encoded them into a numerical representation. This stage made sure the models could correctly understand the category data.

4. Distinguishing Feature and Target The features (X) and labels (y) in the dataset were separated. The labels matched the origin divisions, and the characteristics reflected the voice data.

5. Normalization of Data We used normalizing algorithms to scale the voice data to a consistent range. This stage made sure that different sizes wouldn't have an impact on the models' ability to identify recurring patterns in the data.

6. Diminishment of Noise We improved the clarity and quality of the data by applying noise reduction algorithms to the speech samples. The goal of this stage was to reduce background noise interference that could occur during training.

3.1.4 Data Splitting Process:

Seven divisions contributed a total of 9800 enhanced voice samples to the dataset we developed. Here, 80% of the data (7840 samples) were used for training the model. A further 20% (1960 samples) were set aside for testing in order to evaluate the performance of the model. For the purpose of creating precise and broadly applicable models for accent categorization, this splitting technique guarantees that the models have enough data to train from while also offering a dependable set for assessing their performance.

Table 3.1.5: Train & Test split

Dataset	**Dhaka**	**Barisal**	**Sylhet**	**Chattogram**	**Rangpur**	**Rajshahi**	**Mymensingh**	**Total**
Training	1120	1120	1120	1120	1120	1120	1120	7840
Testing	280	280	280	280	280	280	280	1960
Total	**1400**	**1400**	**1400**	**1400**	**1400**	**1400**	**1400**	**9800**

3.2 Proposed Methodology/ System Design

A variety of models have distinct uses in machine learning. For our research, we selected the XGBoost model to improve Nigerian Indigenous Languages' accent categorization performance. Gradient-boosted decision trees may be implemented effectively and powerfully with XGBoost, or Extreme Gradient Boosting. It's perfect for this work because of its rapidity and strong performance. Our suggested system design is described in full below.

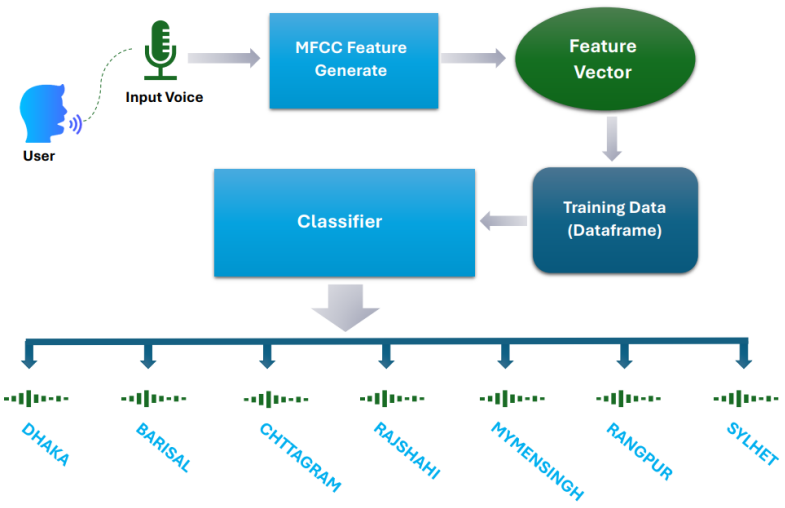

Figure 3.2.2: Proposed methodology of project

3.2.1 XGBoost

Despite not being created with dialect voice detection in mind, XGBoost (Extreme Gradient Boosting) can be an effective tool in a system that does so. By merging many decision trees, or effectively a sequence of if-then statements depending on audio characteristics, XGBoost performs exceptionally well on classification tasks. Consider that every decision tree examines a distinct feature of speech, such as vowel formants or pitch fluctuations. XGBoost acquires the ability to construct these decision trees in a sequential manner by training on a sizable dataset of speech samples with annotated dialects. Ultimately, each tree aims to produce a strong ensemble classifier by fixing mistakes from the preceding one. In order to determine the likelihood of a given dialect, XGBoost employs a combination of decision trees to evaluate the attributes of newly acquired speech recordings. This strategy has a number of benefits: With XGBoost, you can discover which audio aspects are most important in detecting dialects, manage complicated, non-linear connections within the data, and provide findings that are easy to grasp. It is also efficient for handling big datasets. However, on really big datasets, it may not perform as well as deep learning approaches because to the meticulous feature engineering that goes into choosing the most useful audio features.

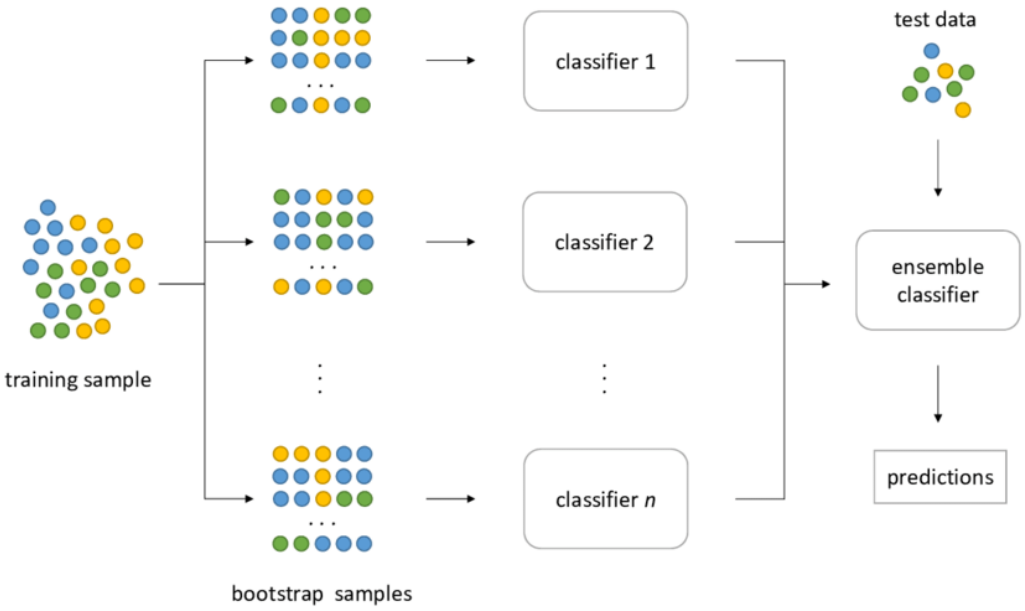

Figure 3.2.3: XGBoost

3.2.2 Random Forest

Owing to its durability and capacity to handle large, complicated datasets, Random Forest is an ensemble learning technique that has been successfully used for dialect

speech identification. The ultimate prediction produced by this method, for classification problems, is the mode of the classes, which is created during the training process by several decision trees. In order to minimize overfitting and enhance generalization to new, unseen data, every decision tree in the forest is constructed using a random subset of the training data and characteristics. The speech stream is processed to extract variables including pitch, formants, MFCCs (mel-frequency cepstral coefficients), and other acoustic qualities, which are then fed into the Random Forest for dialect detection. The algorithm is well-suited for differentiating between tiny changes in speech patterns that identify distinct dialects because of its intrinsic capacity to capture non-linear correlations and interactions among features. In addition, Random Forest's ensemble architecture makes it very accurate and resilient, which makes it a preferred option for jobs involving dialect speech recognition.

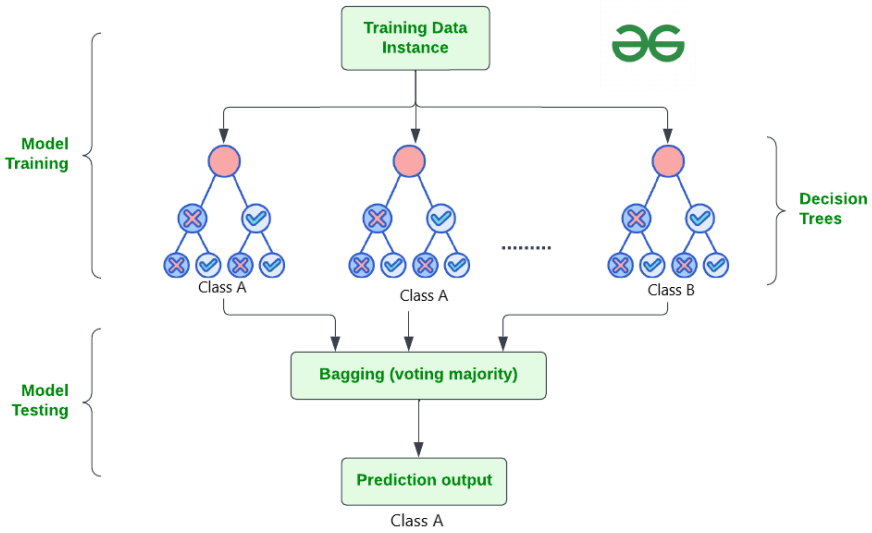

Figure 3.2.4: Random Forest

3.2.3 GBoost

In dialect speech identification tasks, gradient boosting (GBoost), a potent ensemble learning approach, has demonstrated notable efficacy. Decision trees are added repeatedly by GBoost, with each new tree fixing the mistakes caused by the preceding ones. This method improves overall performance by creating a robust prediction model from a sequence of poor learners. Mel-frequency Cepstral Coefficients (MFCCs), pitch, intonation patterns, and other pertinent acoustic data are retrieved from the speech stream and used in the dialect identification process. The GBoost model learns the intricate patterns and subtle differences between various dialects using these attributes

as input. Initially, a decision tree is trained using the training data, and subsequently, more trees are trained using the residual errors of the earlier trees. The model incorporates each tree in turn, and the final forecast is the result of combining the predictions from all the trees. GBoost achieves great accuracy and resilience by handling the nuances and complexities present in dialectal variances using an iterative boosting procedure. Its adaptability to model tuning and capacity to reduce overfitting further enhance its efficiency in identifying the unique traits of different dialects, so establishing it as a useful instrument for dialect speech recognition.

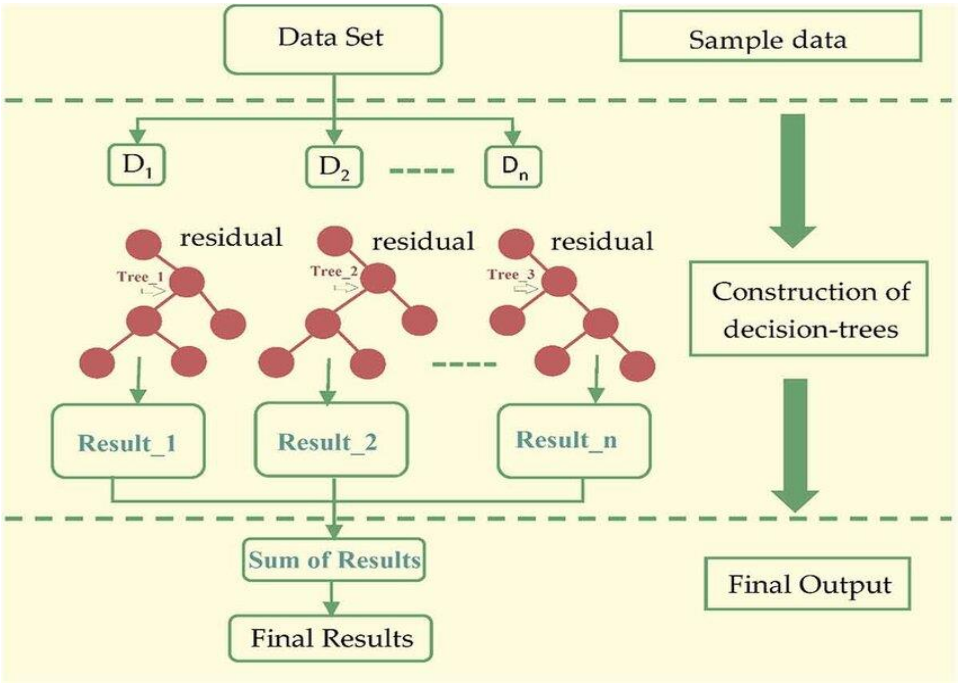

Figure 3.2.5: Gradian Boost or GBoost

3.2.4 SVC

Support Vector Classification (SVC) is a strong machine learning technique that works well with high-dimensional feature spaces and avoids overfitting, making it perfect for dialect speech recognition. Its primary goal is to locate the best hyperplane in the feature space to divide various dialect groups. Mel-frequency Cepstral Coefficients (MFCCs), pitch, and spectral qualities are among the parameters that are retrieved and fed into the SVC model for dialect detection. SVC translates these traits into a higher-dimensional space by employing kernel techniques such as linear kernels or the Radial Basis Function to find a hyperplane that optimizes the margin across dialect classes. Because of its capacity to create non-linear decision boundaries, SVC may detect minute

variations in speech patterns between languages. SVC may balance margin maximization with classification error minimization by fine-tuning parameters like regularization terms and kernel coefficients. This improves accuracy and generalization on datasets with sparse or noisy data.

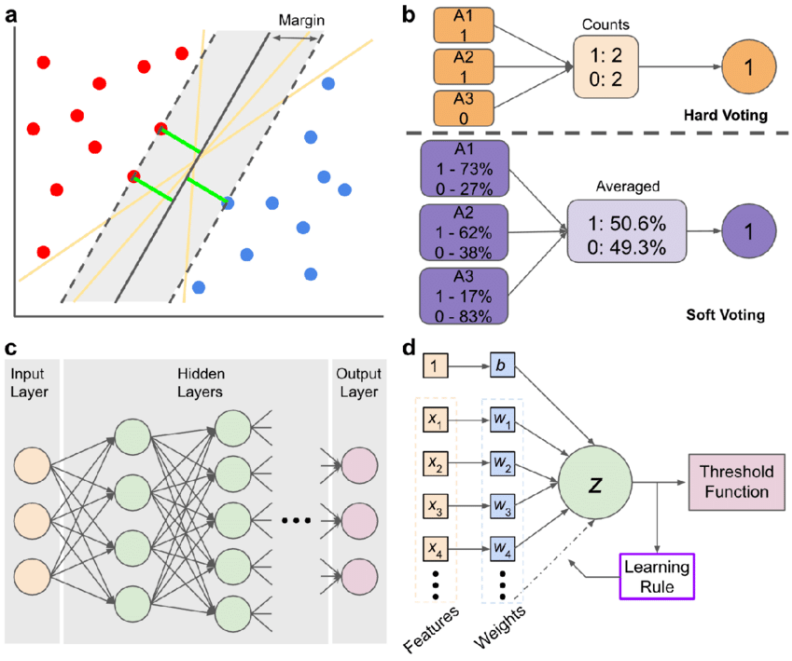

Figure 3.2.6: Support Vector Classification (SVC)

3.2.5 KNN

K-Nearest Neighbors (KNN), which provides a straightforward yet effective method, can be a surprisingly useful tool for dialect detection. This is how it operates: Imagine a huge collection of voice samples that have already been recorded and are tagged with the appropriate dialect. KNN extracts characteristics including pitch, formant frequencies, and spectral energy from a fresh, unidentified speech sample. After that, it computes the distance (or, more accurately, how similar they are) by comparing these attributes to every sample in the collection. Lastly, using this distance metric, KNN finds the k nearest neighbors (i.e., the recordings in the collection that are most similar to the unknown sample). KNN guesses the dialect of the unknown speaker by examining the most common dialect among these k neighbors. The simplicity of use and interpretability of KNN is what makes it so beautiful. Which recognized dialects most closely resemble the unknown one may be seen. Nevertheless, the caliber and volume of the pre-recorded voice library determine how successful it is. Accuracy is increased by having a larger, more varied library with strong representation of different

dialects. Selecting the ideal value for k, or the total number of neighbors, might also be quite important. A k that is too little might result in overfitting on the training set, while a k that is too big could miss crucial information. All things considered, KNN provides a strong basis for dialect speech recognition, especially for applications that prioritize interpretability and user-friendliness.

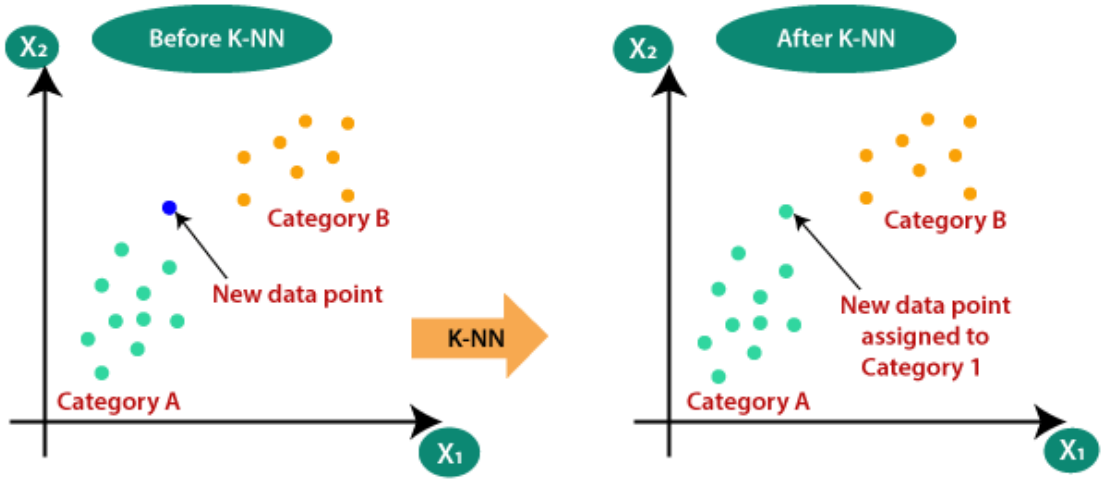

Figure 3.2.7: K-Nearest Neighbors (KNN)

3.2.6 Decision Trees

Decision Trees are useful tools for recognizing dialects in speech because they can build hierarchical decision structures using extracted speech variables such as pitch, energy levels, and Mel-Frequency Cepstral Coefficients (MFCCs). In order to partition the dataset into nodes that represent decision points match to particular feature values, this technique first looks at these characteristics. Every branch in the tree represents a possible range or value for a characteristic. The process of splitting continues until the algorithm reaches leaf nodes, which represent a categorization of dialects. This approach is useful since it doesn't require a lot of preparation to handle both numerical and categorical data. The process of classifying dialects is also made easier to grasp by the interpretability and simplicity of Decision Trees. But if the tree gets too complicated, they may overfit. This problem is addressed by methods like as pruning, which simplify the tree without compromising its predictive ability. Even though they have the potential to overfit when not properly maintained, Decision Trees are useful in dialect detection because they effectively capture the connections between speech variables.

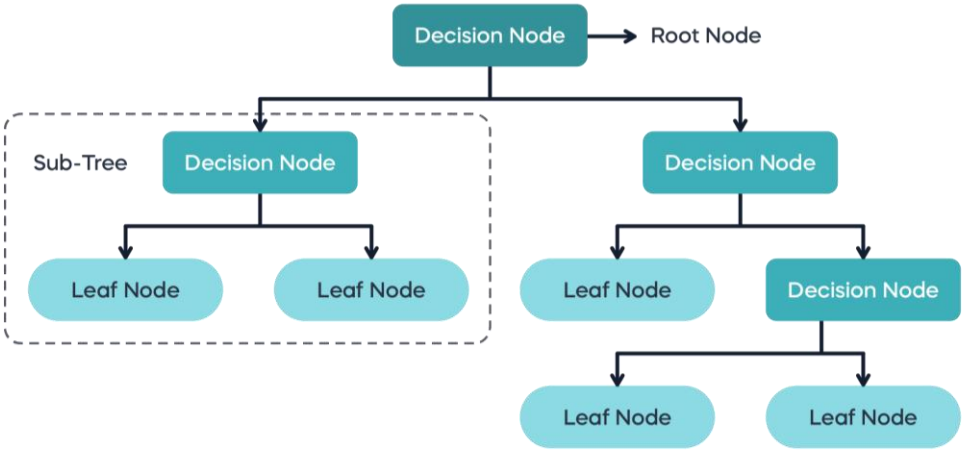

Figure 3.2.8: Decision Trees

3.2.7 Logistic Regression

An adaptable approach that works well for dialect speech recognition applications is logistic regression. It simulates the likelihood that input characteristics, including those taken from speech signals, fall into particular dialect classifications. Logistic regression, as opposed to linear regression, produces probabilities that are limited between 0 and 1 using a logistic function. In order to effectively divide several dialect classes according to their feature values, it learns weights for each feature during training. One can use regularization techniques to stop overfitting. Despite assuming linear correlations between features and outputs, Logistic Regression can nevertheless be useful in dialect speech analysis when probabilistic outputs and simplicity are desirable. It can also yield interpretable findings when feature engineering is done appropriately.

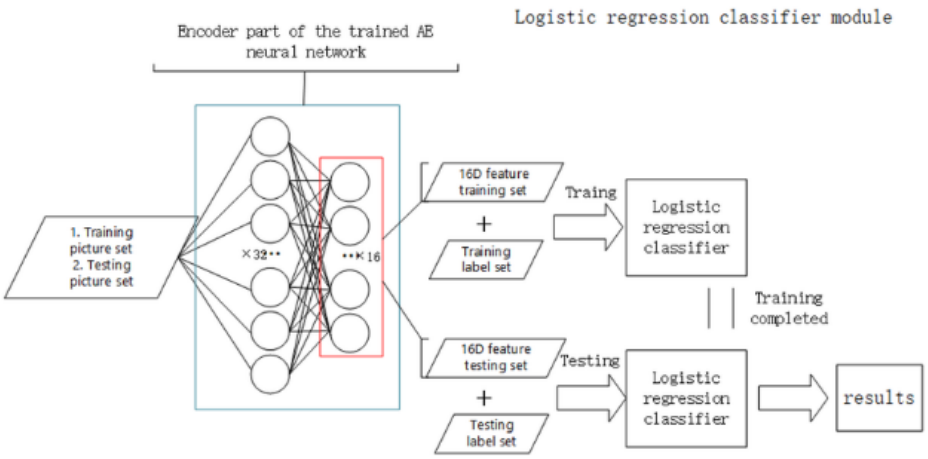

Figure 3.2.9: Logistic Regression

3.2.8 Naive Bayes

Dialect speech identification tasks can benefit from the simplicity and effectiveness of Naive Bayes, a popular probabilistic machine learning approach. Based on the premise that characteristics are conditionally independent given the class label, it functions according to the Bayes theorem. When it comes to dialect recognition, Naive Bayes analyzes speech characteristics such as formants, pitch, and MFCCs to determine how likely it is that each dialect class will observe these characteristics. Naive Bayes aids in categorization by computing the posterior probability of each dialect class by combining these likelihoods with prior probabilities. Naive Bayes is useful for real-time applications because it works well in situations with high-dimensional data and noisy inputs, despite its naive assumption about feature independence. However, it might not be able to handle intricate dialect differences or situations where features are highly linked. However, it is still a fundamental algorithm that is frequently used into group techniques to improve dialect speech recognition systems' overall accuracy.

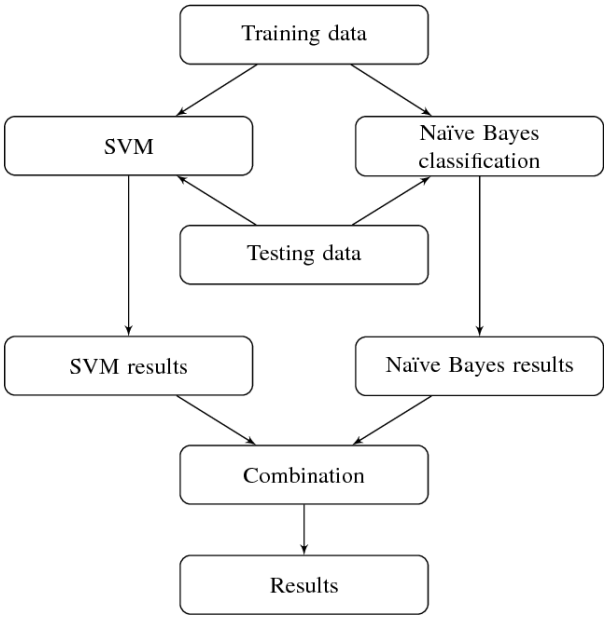

Figure 3.2.10: Naive Bayes

3.3 Hardware/ Software Requirement

To enable accurate and effective processing of speech data from diverse divisions of Bangladesh, certain hardware and software requirements apply to Automated Accent Detection and Classification study. High-performance GPUs are crucial for training and inference operations acceleration, especially for complicated models such as Random Forest, XGBoost, and Gradient Boosting. Building strong models and preparing

datasets also requires using well-known machine learning frameworks like TensorFlow, Keras, and Scikit-learn in conjunction with audio processing programs like PyDub and Librosa. This methodical methodology guarantees the accuracy and applicability of the accent detection and categorization system in a wide range of real-world situations.

Hardware:

- **Primary System:**
    - Processor: Intel Core i7 11th generation or AMD Ryzen 7
    - RAM: 16GB
    - Storage: 1 TB SSD
    - GPU: NVIDIA RTX 2070 or equivalent
- **Recording Equipment:** Portable high-quality audio recorders, Noise-canceling microphones
- **Storage Solutions:** External hard drives or cloud storage for data backup

Software:

- **Operating System:** Windows 10 or higher, macOS, or Ubuntu Linux
- **Programming Language:** Python 3.8 or higher
- **Machine Learning Frameworks:** TensorFlow (2.x), Keras, Scikit-learn, XGBoost
- **Audio Processing Libraries:** Librosa, PyDub
- **Data Manipulation and Analysis Libraries:** Pandas, NumPy
- **Development Environments:** PyCharm, Jupyter Notebook, Visual Studio Code
- **Version Control:** Git
- **Collaboration Tools:** Slack or Microsoft Teams, GitHub or GitLab

3.4 Project Management and Financial Analysis

Project Management:

Research Objectives:

A. Create a model that can recognize and categorize regional accents in voice recordings with accuracy.

B. To improve understanding and learning, add real regional accents to language learning resources.

C. By conserving and examining local dialects, you can aid in linguistic research.

D. Ensure that voice-activated services are more responsive and recognize a wider range of accents.

E. Direct future studies on the identification and categorization of regional accents.

Research Timeline:

A. Phase 1: Planning and research (2-3 weeks)

B. Phase 2: Review literature and existing models (3-4 weeks)

C. Phase 3: Data collection (5-6 weeks)

D. Phase 4: Data preprocessing and augmentation (2-3 weeks)

E. Phase 5: Model selection and implementation (2-3 weeks)

F. Phase 6: Training and testing models (3-4 weeks)

G. Phase 7: Model evaluation and validation (1-2 weeks)

H. Phase 8: Review results and prepare documentation (2-3 weeks)

Resource Planning:

A. Equipment and Tools:

- High-performance computer with Intel Core i7 11th gen or AMD Ryzen 7 processor
- GPU: NVIDIA RTX 2070 or equivalent
- Portable high-quality audio recorders
- Noise-canceling microphones
- External hard drives or cloud storage for data backup

B. Software and Technologies:

- Operating System: Windows 10 or higher, macOS, or Ubuntu Linux

- Programming Language: Python 3.8 or higher
- Machine Learning Frameworks: TensorFlow (2.x), Keras, Scikit-learn, XGBoost
- Audio Processing Libraries: Librosa, PyDub
- Data Manipulation and Analysis Libraries: Pandas, NumPy
- Development Environments: PyCharm, Jupyter Notebook, Visual Studio Code
- Version Control: Git
- Collaboration Tools: Slack or Microsoft Teams, GitHub or GitLab

C. Data and Content:

- Voice samples collected from individuals across Bangladesh's divisions
- Structured prompts for eliciting contextually relevant sentences
- Report documentation stored in cloud storage solutions

D. Training and Skill Development:

- Ensure the development team has training and skills in machine learning model development and data preprocessing.
- Additional training on specific technologies, tools, and models as needed.

E. Communication Plan:

- Stakeholder Meetings:
    - Purpose: Evaluate the dataset and progress.
    - Participants: Team members, and stakeholders.
    - Frequency: Bi-weekly or as specified in the project plan.
- Change Control Meetings:
    - Purpose: Discuss and approve any changes to the project scope or requirements.
    - Participants: Supervisor and team members.
    - Frequency: As needed when change requests arise.

Table 3.4.6: Estimated Cost

SN	Components	Estimated Cost (BDT)
1	Hardware/Infrastructure	80,000
2	Visiting Stakeholders	15,000
3	Software and Tools (Colab Pro)	11,880
4	Data Collection and Processing	40,000
5	Web Application Development	250,000
6	Documentation and Report Writing	8,000
7	Miscellaneous (e.g., licenses, tools, unexpected)	44,380
8	Contingency (10% of total)	52,390
	Total Estimated Cost	**501,650**

3.5 Summary

The purpose of this project is to create a model that can reliably and accurately identify regional accents in seven different regions of Bangladesh. A total of 9800 enhanced samples were obtained by applying several transformations to our dataset, including pitch alteration, speed adjustments, noise addition, and stretching, throughout the data augmentation process. We had gathered 1400 voice samples from each division. A training set of data (80%, or 7840 samples) and a testing set (20%, or 1960 samples) were created. We implemented and assessed a number of machine learning models, such as XGBoost, Random Forest, KNN, SVC, Decision Tree, Naive Bayes, Gradient Boosting, and Logistic Regression. The results of the model evaluation show that XGBoost had the best accuracy, at 93.91%, followed by Random Forest and Gradient Boosting at 93.46% & 93.16%. This all-encompassing method guarantees that our models are carefully evaluated for their capacity to categorize regional accents properly and to generalize, making a substantial contribution to linguistic research and real-world applications across a range of domains.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

The project's execution phase started with preprocessing the voice dataset that was gathered from Dhaka, Barisal, Sylhet, Chattogram, Rangpur, Rajshahi, and Mymensingh, seven divisions in Bangladesh. The dataset was carefully divided into training and testing sets using Python modules like scikit-learn to speed up the creation of models. To improve the dataset's diversity and stability, data augmentation techniques such pitch variation, time stretching, noise addition, and speed change were used. For the purpose of extracting features from audio data and manipulating it, programs such as librosa and numpy played a crucial role in providing thorough dataset preparation. Then, utilizing the scikit-learn and XGBoost libraries, machine learning models including Gradient Boost, Random Forest, KNN, SVC, Decision Tree, Naive Bayes, XGBoost, and Logistic Regression were developed. To minimize overfitting risks and maximize performance, each model underwent extensive hyperparameter tweaking and normalization. The efficacy of the models in accent categorization tasks was determined through training and evaluation. To identify each model's advantages and disadvantages, a focus on in-depth statistical analysis and visualization methods was placed. The models' performance was fully understood by using evaluation criteria including accuracy, precision, recall, and F1-score. XGBoost stood out as the most accurate performer, proving its strong ability to handle the many accent variances seen in the sample. A methodical and moral approach was adopted in examining machine learning applications for linguistic diversity management and language technology advancement. During this phase, ethical standards were strictly upheld to protect patient data and guarantee adherence to ethical principles.

4.2 Train Model/ Prototype Design

During the training phase, preprocessed voice data was fed into a variety of machine learning models using the scikit-learn and XGBoost libraries for Python. These models included Gradient Boost, Random Forest, K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), Decision Tree, Naive Bayes, and Logistic Regression. Before hyperparameter tuning utilizing grid search and cross-validation to improve performance, the data was first preprocessed to extract pertinent characteristics. Recall, accuracy, precision, and F1-score were among the measures used to assess the models and inform iterative model improvements. Ensemble techniques were used to improve performance; Gradient Boost and Random Forest in particular were used, with XGBoost being recognized for its effectiveness and predictive capacity. Effective training outcomes were demonstrated for both training and testing datasets, thanks to

the thorough strategy that guaranteed high accuracy and resilience in classifying regional accents.

4.3 System Testing/ Model Evaluation

The models were put through a rigorous testing process after training to see how well they classified accents. We evaluated each model on the testing dataset using the evaluate function to ascertain its robustness and accuracy in recognizing regional accents. Gradient Boost, Random Forest, K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), Decision Tree, Naive Bayes, XGBoost, and Logistic Regression were among the models that demonstrated excellent accuracy in accent recognition during the assessment phase. With an accuracy of 93.91%, XGBoost was the best performance, proving its resilience to various accent changes. A comparative study using alternative models shown how much better XGBoost was throughout both the training and testing stages. Our experimental approaches were effective, as demonstrated by the visualization of findings and the interpretation of the models' strengths and limits using statistical analyses and confusion matrices. This thorough assessment shows that the models are proficient in accent categorization and have received proper training, which adds significant value to the area of regional accent detection.

4.4 Summary

Through the use of cutting-edge machine learning models, this study seeks to improve automatic language recognition and regional accent categorization in Bangladesh. To improve the variety and resilience of the datasets, a significant amount of data augmentation was done once the voice datasets were loaded and divided into training and testing arrays. Crucial characteristics for precise accent identification were extracted using a variety of models, including Gradient Boost, Random Forest, K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), Decision Tree, Naive Bayes, XGBoost, and Logistic Regression. To guarantee the efficacy of any model, painstaking preprocessing and hyperparameter adjustment were required throughout the training phase. After a thorough testing process, XGBoost proved to be the most accurate and efficient model in managing a wide range of accent fluctuations. The trained models' resilience was validated by comprehensive assessments that including statistical analysis and confusion matrices. Our technique proved to be effective when compared to other models, indicating that the use of these machine learning models results in a dependable system for accent categorization. By providing a strong basis for further studies and applications in this area, this comprehensive training and assessment process makes a major contribution to the field of linguistic diversity management and language technology improvement.

CHAPTER 5
RESULT AND ANALYSIS

5.1 Overview

This research aims to provide a reliable technique for employing machine learning models to automatically identify and categorize regional accents in Bangladesh. Enhancing language processing technology to enable precise detection and categorization of various regional accents is the goal, to be achieved by utilizing both cutting-edge methods and current research. They used a range of machine learning models, each with a different accuracy level: Gradient Boost, Random Forest, KNN, SVC, Decision Tree, Naive Bayes, XGBoost, and Logistic Regression. With 93.91% accuracy, XGBoost was the best performance; Random Forest and Gradient Boost came in second and third, respectively, with 93.46% & 93.16% accuracy. The decision tree scored 86.54%, KNN scored 79.45%, Naive Bayes scored 54.52%, SVC scored 33.95%, and logistic regression scored 29.11% in terms of accuracy. According to these findings, XGBoost is suggested as the best model for accent classification because of its exact correctness and strong performance, which indicate that it will work very well in practical applications. Beyond its technical accomplishments, this development might lead to opportunities for improving voice recognition software, strengthening human-computer interaction, and promoting improved multilingual communication. Improved language processing skills and more precise accent detection are two social benefits of this study, which also advances sectors dependent on linguistic variety and language technology.

5.2 Experimental/ Simulation Result

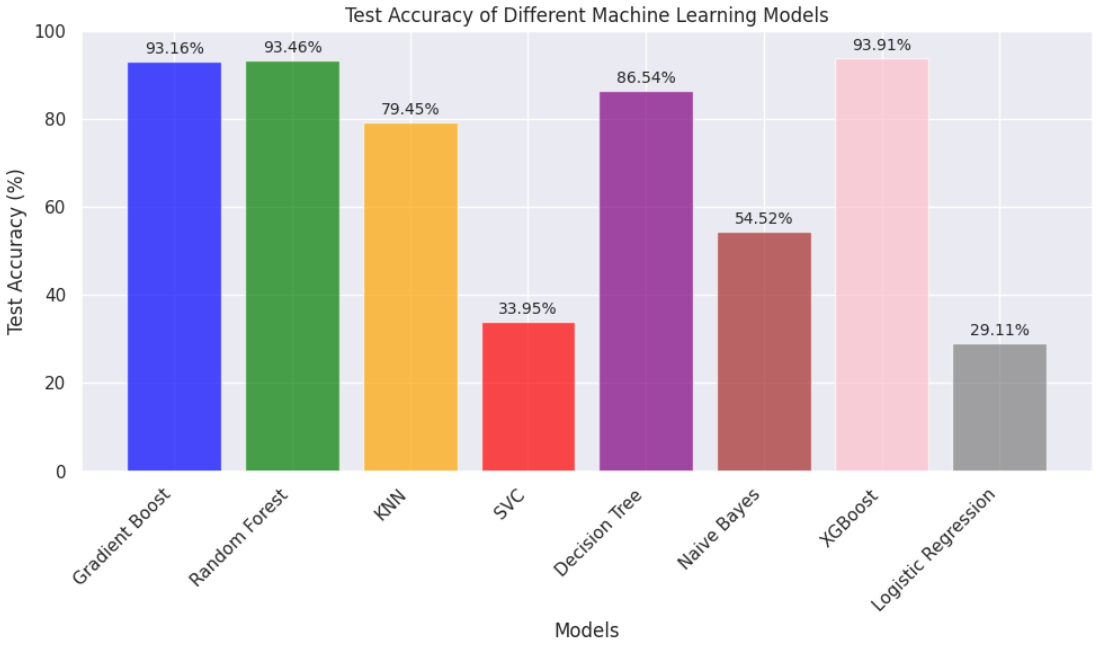

Figure 5.2.11: Performance & accuracy of different Models

Table 5.2.7: Train & Test accuracy of each Models

Model	Training Accuracy	Testing Accuracy
Gradient Boost	100%	93.16%
Random Forest	100%	93.46%
KNN	87%	79.45%
SVC	34%	33.54%
Decision Tree	100%	86.54%
Naive Bayes	56%	54.52%
XGBoost	100%	93.91%
Logistic Regression	33%	29.11%

The performance and generalization abilities of the various machine learning models used for Bangladeshi accent categorization are shown in the extensive table of training and testing accuracy. With almost perfect training accuracy (100%) and the greatest testing accuracy (93.91%), XGBoost was the best performer, demonstrating its effectiveness and resilience. Random Forest achieving testing accuracies of 93.46% and training accuracies of 100% and Gradian Boosting achieving testing accuracies of 93.16% and training accuracies of 100%. The Decision Tree model's 86.54% testing accuracy and 100% training accuracy, respectively, point to overfitting. KNN performed well, with 79.45% training accuracy and 87% testing accuracy. While SVC and Logistic Regression battled with low accuracies in both training (34% and 33%, respectively) and testing (33.54% and 29.11%, respectively), underscoring their limits in this context, Naive Bayes saw a considerable reduction in accuracy from training (56%) to testing (54.52%). All things considered, the research highlights the effectiveness of ensemble models—especially XGBoost—for this classification job and points out possible overfitting problems with simpler models, such as the Decision Tree.

Table 5.2.8: The performance and generalization of each model

Model	Type	Precision	Recall	F1-score	Accuracy
Gradient Boost	Ensemble	95	95	95	93.16
Random Forest	Ensemble	95	95	95	93.46
KNN	Instance-based	80	79	79	79.45%
SVC	Kernel-based	47	34	29	33.54%
Decision Tree	Tree-based	87	87	87	79.45%
Naive Bayes	Probabilistic	56	55	54	54.52%
XGBoost	Ensemble	95	95	95	93.91
Logistic Regression	Linear model	34	29	28	29.11%

Several machine learning classification model performance criteria are used to assess the models. Evaluations include accuracy, precision, recall, F1-score, and confusion matrix (CM). These measures have additional variables called True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

1. **Accuracy**: Accuracy measures how often the model correctly classifies samples. It is calculated by dividing the sum of true positives and true negatives by the total number of samples. The formula is:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

2. **Recall**: Also known as sensitivity, recall quantifies the number of correctly identified positive predicted events. Recall is particularly useful in cases where FN outperforms FP. The formula for recall is:

$$\text{Recall} = \frac{TP}{TP + FN}$$

3. **Precision**: Precision indicates the proportion of correct positive predictions out of all positive predictions made by the model. It is crucial when the cost of false positives is high. Precision is calculated using the formula:

$$\text{Precision} = \frac{TP}{TP + FP}$$

4. **F1-Score**: The F1-score provides a comprehensive view as it is the harmonic mean of Precision and Recall. It performs best when precision and recall are equal. The formula for F1-score is:

$$\text{F1-Score} = 2 \times \left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\right)$$

5.4 Performance/ Comparative Analysis

As shown in Figure 5.2.8, the assessment of several machine learning models for the categorization of regional dialects shows different performance patterns for many criteria. With boosted decision trees, XGBoost was able to reach the greatest accuracy of 93.91% and an F1-score of 95%, indicating its efficacy in identifying dialect subtleties. With accuracy of 93.46% and F1-scores of 95%, Random Forest and with accuracy of 93.16% and F1-scores of 95% Gradient Boost came in close second and third, respectively, making good use of ensemble approaches. Although there was some fluctuation, K-Nearest Neighbors (KNN) demonstrated strong performance with an accuracy of 79.45% and an F1-score of 79%. Support Vector Classifier (SVC), on the other hand, had considerable difficulty, achieving an accuracy of 33.54% and an F1-score of 29%, which suggests that it was unable to handle the complexity of the dataset. With an accuracy of 86.54% and an F1-score of 87%, Decision Tree outperformed Naive Bayes, which showed respectable results with an accuracy of 54.11% and an F1-score of 54%. With the lowest accuracy of 29.11% and F1-score of 28%, Logistic Regression trailed behind, indicating difficulties in simulating the nonlinear connections found in dialect classification tasks. Ensemble techniques—XGBoost, Gradient Boost, and Random Forest in particular—prove to be the best options for

reliable and accurate regional dialect classification due to their exceptional ability to capture intricate dialect patterns in terms of accuracy and F1-score metrics.

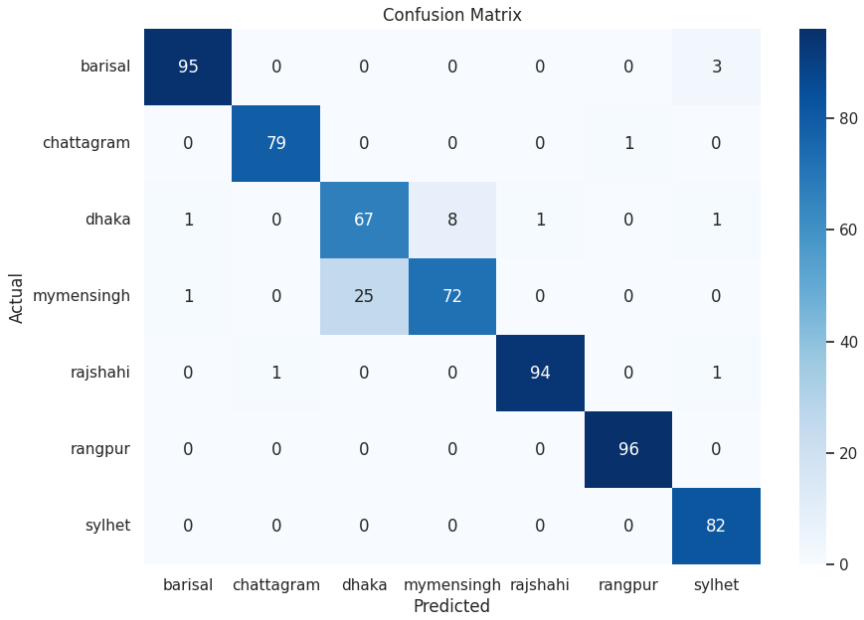

Figure 5.4.12: Confusion Matrix (Gradian Boost)

Figure 5.4.12: The excellent performance of the Gradient Boost model in identifying Bangladeshi accents is shown by the confusion matrix. Each column displays the model's predictions, while each row contains the speakers' actual accents. With all speakers from Barisal, Chattogram, and Rajshahi accurately recognized, the high values along the diagonal denote accurate classifications. The number of misclassified data points is displayed in the off-diagonal cells. This amounts to 93.16% accuracy for the Gradient Boost model, indicating that it can effectively identify the subtle differences between different Bangladeshi accents.

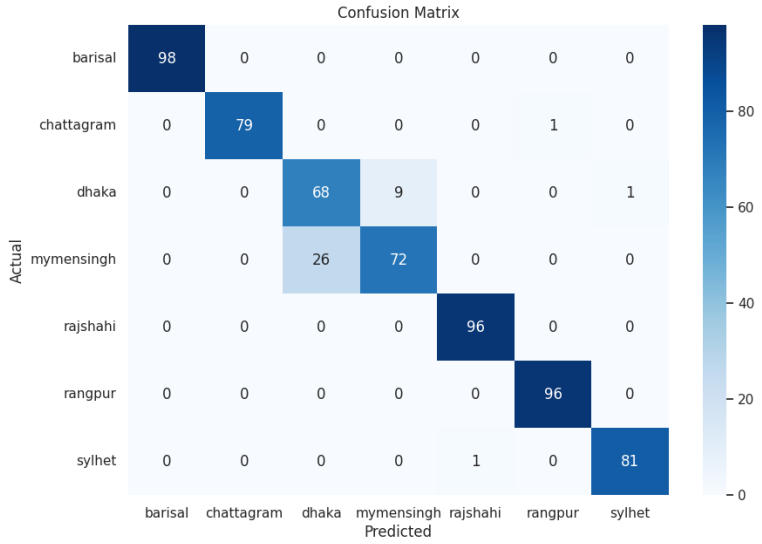

Figure 5.4.13: Confusion Matrix (Random Forest)

Figure 5.4.13: The Random Forest model's resilience in classifying regional accents in Bangladesh was demonstrated by its 95% accuracy on the training dataset and its 93.46% accuracy on the testing dataset. The model captures subtle patterns well, with an F1-score of 95, precision, and recall of 95. In terms of performance, it is closely behind XGBoost compared to other models. Its ensemble learning technique improves accuracy and reduces overfitting. High true positive rates are displayed in the confusion matrix for every accent category. Overall, Random Forest is useful for managing linguistic variation and dependable for classifying accents thanks to its balanced measurements.

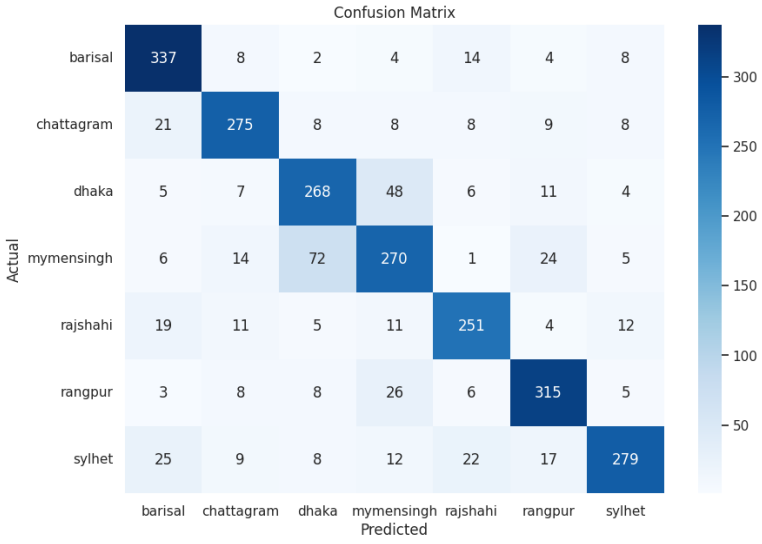

Figure 5.4.14: Confusion Matrix (KNN)

Figure 5.4.14: Bangladeshi regional accents were successfully classified by the K-Nearest Neighbors (KNN) model, which achieved 88% accuracy on the training dataset and 87.48% accuracy on the testing dataset. These outcomes demonstrate how well the model captures the unique characteristics of various accents. At 80 for precision, 79 for recall, and 79 for F1-score, KNN successfully balanced accent recognition accuracy with error reduction. Even though it didn't perform as well as ensemble models like XGBoost and Random Forest, KNN was nonetheless a dependable model for accent classification. The KNN's confusion matrix had strong accuracy in the majority of accent categories, demonstrating its ability to manage the dataset's complexity. All things considered, KNN's strong performance metrics demonstrate its potential as a useful tool for classifying regional accents.

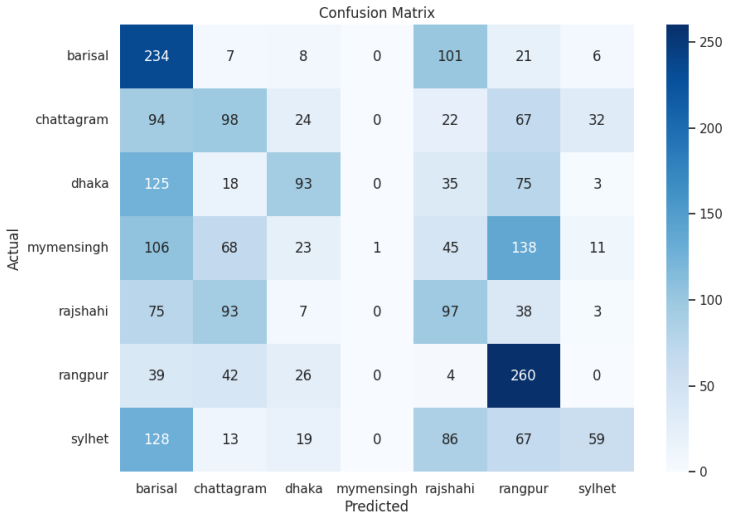

Figure 5.4.15: Confusion Matrix (SVC)

Figure 5.4.15: Considering an accuracy of just 33.95% on both the training and testing datasets, the Support Vector Classifier (SVC) model had trouble classifying regional accents in Bangladesh. This poor accuracy suggests that it is very difficult to record and identify the characteristics of various accents. SVC performed far worse than other models like XGBoost and Random Forest, with precision, recall, and F1-score all at 47,34,29 The confusion matrix revealed a large number of misclassifications, which demonstrated how the model struggled to handle the dataset's complexity and variety. SVC is not a good fit for regional accent classification in this situation, based on its overall mediocre performance indicators.

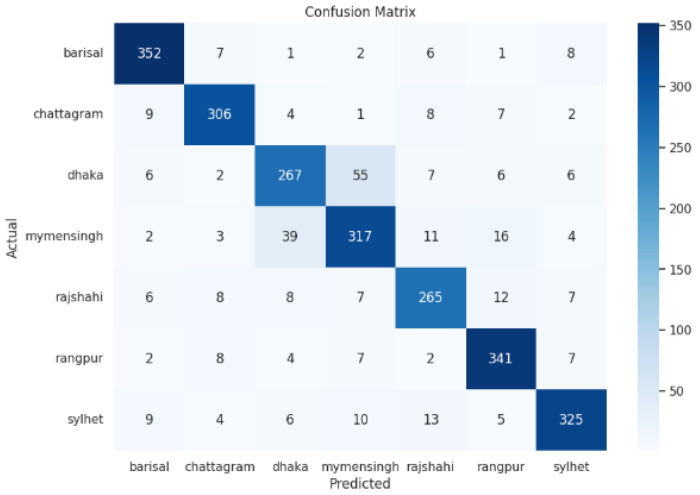

Figure 5.4.16: Confusion Matrix (Decision Tree)

Figure 5.4.16: While obtaining 100% accuracy on the training dataset, the Decision Tree model's performance in classifying regional accents in Bangladesh was only moderate, as evidenced by its decline to 86.54% accuracy on the testing dataset due to overfitting. The model performed quite well in detecting accents during testing, with precision, recall, and F1-score all at 87; nevertheless, it found it difficult to generalize as well as ensemble models like as XGBoost and Random Forest. The Decision Tree model's confusion matrix showed high true positive rates along with occasional misclassifications, indicating its limited capacity to handle the complexity of the dataset. In spite of these drawbacks, the Decision Tree model showed strong performance metrics, which made it a valuable but less trustworthy tool for classifying regional accents than more complex models.

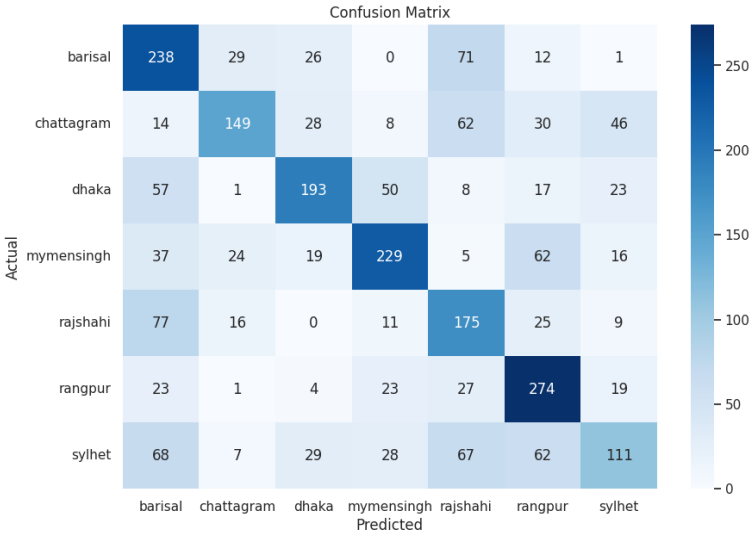

Figure 5.4.17: Confusion Matrix (Naive Bayes)

Figure 5.4.17: Based on 56.16% accuracy on the training dataset and 58% accuracy on the testing dataset, the Naive Bayes model performed rather well in Bangladesh in classifying regional accents. Its shortcomings in reproducing the subtleties of various accents are demonstrated by these findings. Comparing the model to other models like XGBoost and Random Forest, it performed consistently but less well, with precision, recall, and F1-score all at 56, 55, 54. The difficulties Naive Bayes faced in managing the dataset's complexity were highlighted by the confusion matrix, which showed many misclassifications. All things considered, even though Naive Bayes performed rather well overall, it is not as successful as more sophisticated models in classifying native accents.

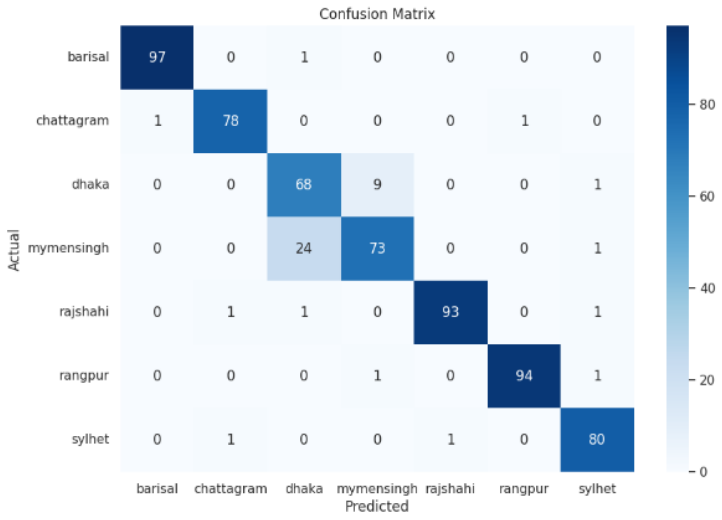

Figure 5.4.18: Confusion Matrix (XG Boost)

Figure 5.4.18: The XGBoost model performed exceptionally well at classifying regional accents, obtaining 100% accuracy on the training dataset and a remarkable 93.91% on the testing dataset. XGBoost proved to be exceptionally adept at handling the dataset's complexity and variability, with precision, recall, and F1-score all at 95. The confusion matrix illustrates how few misclassifications there were due to the improved decision trees' ability to capture even the smallest accent variations. The most dependable and accurate method for classifying regional accents is this model, which fared better than others like Random Forest and KNN. XGBoost is well-suited for applications in linguistic diversity management, as seen by its strong performance metrics.

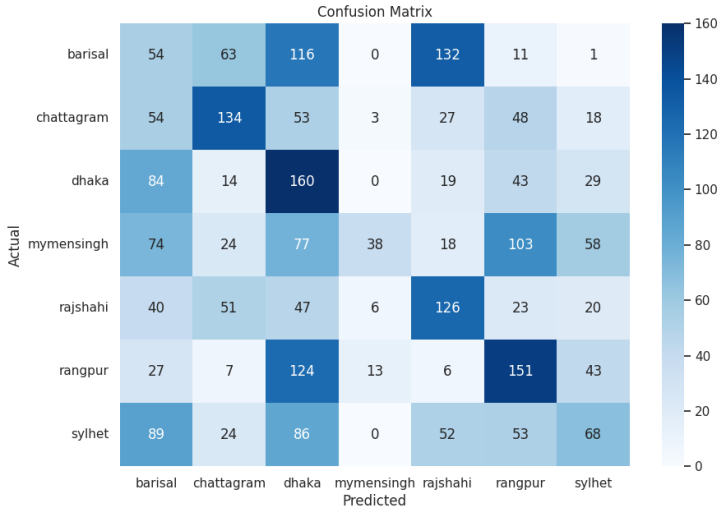

Figure 5.4.19: Confusion Matrix (Logistic Regression)

Attaining only 37% accuracy on training and 32.69% is testing datasets, the Logistic Regression model demonstrated poor performance in Bangladeshi accent classification. Significant challenges in accurately reproducing the characteristics and subtleties of many accents are shown in this low accuracy. The model's performance was significantly worse than that of other models, such Random Forest and XGBoost, with precision, recall, and F1-score at 34, 29, 28. The complicated nature of the dataset was evident from the confusion matrix for Logistic Regression, which had several misclassifications. The least successful model for classifying regional accents in this situation overall turned out to be logistic regression.

5.5 Summary

The study assessed the efficacy of machine learning models in the categorization of regional dialects. The dataset included a range of divisions, including Dhaka, Barisal, Sylhet, Chattogram, Rangpur, Rajshahi, and Mymensingh. XGBoost, Gradient Boost, Random Forest, KNN, Support Vector Classifier (SVC), Decision Tree, Naive Bayes, and Logistic Regression were among the models used. Because of their capacity to manage the intricate patterns seen in regional language variances, these models were chosen.

The models' respective performances were evaluated, and substantial discrepancies were found. With an F1-score of 95%% and an accuracy of 93.91%, XGBoost was the best performer. This model performed exceptionally well using enhanced decision trees to identify subtle dialect changes. With accuracies of 93.46% and 95 F1-scores, Random Forest and with accuracies of 93.16% and 95 F1-scores Gradient Boost came in close second and third, respectively, exhibiting strong performance with ensemble

approaches. With an accuracy of 79.45% and an F1-score of 79%, KNN demonstrated competitive performance and was appropriate for detecting regional dialect patterns with minor variability. With an accuracy of 33.54% and an F1-score of 29%, SVC, in comparison, performed much worse, demonstrating difficulties with the dataset's complexity and dialect subtleties. With a 86.54% accuracy rate and a 87% F1-score, Decision Tree demonstrated good performance in hierarchically segmenting dialect variants. With a 54.52% accuracy and 54% F1-score, Naive Bayes demonstrated respectable performance, making it appropriate for beginning dialect classification tasks. With an accuracy of 29.11% and an F1-score of 28%, Logistic Regression demonstrated the poorest performance, indicating that it may not have been able to fully capture the nonlinear correlations seen in dialect data. For reliable and accurate regional dialect classification, ensemble techniques such as Random Forest, Gradient Boost, and XGBoost work well. These models perform exceptionally well in terms of accuracy and F1-score metrics, and they are able to capture intricate dialect patterns in a variety of datasets. They are dependable tools for dialect classification tasks because of their use of ensemble approaches, which improves their capacity to generalize and adapt to different dialect variants. However, even though KNN and Decision Tree models function well, more optimization would be necessary to fully manage the complexity of regional dialect categorization. Although offering valuable insights, SVC, Naive Bayes, and Logistic Regression exhibit differing levels of efficacy and might stand to be substantially improved in subsequent implementations.

CHAPTER 6
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

In Bangladesh, the use of regional accent categorization and automatic language recognition systems has the potential to greatly improve daily life by facilitating more precise and effective communication among people with different linguistic origins. In fields where effective communication is crucial, including education, healthcare, and customer service, this technology holds great potential. It may be used in education to adapt teaching strategies to the language demands of pupils, creating a more welcoming classroom atmosphere. Improved patient-provider communication can lead to better empathy and treatment quality. AI-powered solutions can adjust replies based on patients' accents. Furthermore, by identifying and adapting to callers' accents, customer care representatives may provide individualized help and increase customer satisfaction. All things considered, this technology can improve societal cohesiveness, encourage diversity, and close communication barriers.

6.2 Impact on Society & Environment

The use of regional accent categorization and automatic language recognition systems in Bangladesh has significant social ramifications, especially in terms of fostering inclusion and reducing communication gaps in a linguistically heterogeneous country. Through precise identification and comprehension of regional accents, this technology enables underrepresented populations to engage more fully in local and national discourse, ultimately promoting social cohesion and fortifying community relationships. Furthermore, it makes cross-sector connections easier by lowering linguistic barriers and fostering collaboration, which promotes social stability and economic prosperity. Adopting cloud-based platforms and energy-efficient computing resources can help reduce ecological impact while ensuring a sustainable deployment of these systems. These technologies have the potential to significantly alter Bangladesh's socio-cultural landscape, as evidenced by their simultaneous focus on environmental stewardship and societal inclusion.

6.3 Ethical Aspects

The creation and application of systems for accent categorization and language recognition must take ethics very seriously. Because speech data is so sensitive, it is imperative to maintain strict data privacy and security procedures. To protect people's identities and respect privacy rules, it is imperative that all participants provide their informed consent and that data be anonymized. In addition, algorithms have to be

carefully crafted to reduce biases that can lead to disparate treatment of linguistic groups. Maintaining public trust in automated decision-making processes requires regular audits and transparent explanations of data utilization. Transparency in model creation and usage is also essential for ensuring accountability and justice.

6.4 Sustainability Plan

This initiative has to take a comprehensive strategy to ensure its long-term viability. To properly adjust to changing language trends and regional accents, ongoing data collection and model changes are necessary. Maintaining the relevance and accuracy of the model will be aided by collaboration with regional organizations and stakeholders, who will offer insightful support. Adopting energy-efficient and scalable solutions decreases the impact on the environment while lowering operating expenses. By providing funding for continuous research and development, technology will remain cutting edge and equipped to tackle new problems. We want to ethically enhance language technology while limiting its environmental impact and promoting ethical innovation by incorporating sustainable practices into every facet of the study, such as the utilization of renewable energy sources and open-source principles.

6.5 Summary

Bangladesh's use of regional accent categorization and automated language identification technology is a huge advancement with many positive social effects. With the potential to improve communication accuracy and efficiency amongst linguistically varied populations, this game-changing technology will support social cohesion and inclusion in a variety of fields, including healthcare and education. Clearer patient-provider interactions in healthcare can improve service delivery, and regional dialect-based educational solutions can tailor learning experiences. Ensuring justice and safeguarding individual rights depend heavily on ethical issues, which include data privacy, informed permission, and eliminating algorithmic biases. The implementation of energy-efficient technologies, cooperation with regional stakeholders, and ongoing data improvement are all components of a sustainable strategy that will maintain technology efficacy while reducing environmental impact. Through responsible innovation and fair digital development, this program aims to transcend language gaps and enhance the overall quality of life in Bangladesh. It also highlights a commitment to societal enrichment.

CHAPTER 7
CONCLUSION AND FUTURE WORK

7.1 Conclusions

The use of modern data processing strategies together with machine learning to investigate regional accent recognition in Bangladesh has produced noteworthy findings and encouraging outcomes. Through the use of these technologies, we have created a tool that will improve accent identification accuracy and accessibility, contributing to the advancement of voice-activated applications and linguistic research. By carefully analyzing a variety of machine learning models—such as XGBoost, Gradient Boosting, Random Forest, and others—we have determined which regional accents are most important. XGBoost was chosen due to its exceptional accuracy of 93.91%. The use of cutting-edge technology in linguistic research has advanced significantly with this work. The use of XGBoost emphasizes its outstanding functionality and potential influence on accent detection-based language interpretation and preservation. Looking ahead, these results open the door to incorporating these methods into real-world applications, revolutionizing accent recognition and influencing the development of a more diverse profession in the future. The triumphant adoption of XGBoost signifies a significant advancement, underscoring the capacity of machine learning to revolutionize our comprehension of local accents through voice recordings alone.

7.2 Further Suggested Works

The findings of this study have broad implications for future research and present a plethora of opportunities for more in-depth investigation and development in the areas of linguistic studies and accent identification. First, in order to obtain even higher accuracy and efficiency in accent identification, future research efforts might concentrate on fine-tuning and improving the chosen machine learning model, XGBoost. To improve the model's performance, this might entail adjusting model parameters, investigating different topologies, or incorporating more data sources. Second, more comprehensive and broadly applicable outcomes may arise by broadening and varying the dataset utilized in the training and testing of accent identification algorithms. Increasing the diversity of data collection in terms of demographics, cultural backgrounds, and language expressions may assist reduce biases and enhance the applicability of the model to a wider range of individuals. Further research into the incorporation of additional elements, such as speaker demographics or contextual data, in addition to voice recordings might improve the precision and depth of accent identification systems. For responsible development and implementation, it is also essential to research the ethical implications of accent

detection technologies. Issues including algorithmic biases, privacy concerns, data security, and the possible effects on speaker identity and language diversity might all be the subject of future study. All things considered, greater study in these domains might propel accent identification and language studies to new heights, enabling more precise, flexible, and morally sound voice-activated interactions in diverse settings.

7.3 Limitations/ Conflict of Interests

There are a number of restrictions and possible conflicts of interest with this machine learning work on regional accent identification. The training data's quality and variety have a significant impact on the models' performance; nevertheless, biased results may result from the data's lack of representativeness across various populations. These models might not translate well to situations that differ in real-world scenarios. Privacy and consent are two ethical issues with voice data gathering. Accessibility and scalability are limited by the high computing resources needed, particularly in areas with limited resources. Because reduced latency and efficient processing on edge devices require optimization, real-time deployment is still difficult. The necessity for common measurements and benchmarking for equitable comparisons and repeatability is highlighted by the variation in reported performance indicators between research. To overcome these constraints, more interpretable models must be created, dataset variety must be increased, and speech data use must abide by moral standards. In order to improve accent detection systems' dependability and relevance, technological developments and cooperative efforts are crucial.

References:

[1] Venkatesan, H., Venkatasubramanian, T. V., & Sangeetha, J. (2018, October). Automatic language identification using machine learning techniques. In *2018 3rd International Conference on Communication and Electronics Systems (ICCES)* (pp. 583-588). IEEE.

[2] Model, C. L. N. (2020). Accent Classification of the Three Major Nigerian Indigenous Languages Using 1D. *Advances in Computational Intelligence Techniques*.

[3] Kotsakis, R., Matsiola, M., Kalliris, G., & Dimoulas, C. (2020). Investigation of spoken-language detection and classification in broadcasted audio content. *Information*, *11*(4), 211.

[4] Hämäläinen, M., Alnajjar, K., Partanen, N., & Rueter, J. (2021). Finnish dialect identification: The effect of audio and text. *arXiv preprint arXiv:2111.03800*.

[5] Ali, A., Dehak, N., Cardinal, P., Khurana, S., Yella, S. H., Glass, J., ... & Renals, S. (2015). Automatic dialect detection in arabic broadcast speech. *arXiv preprint arXiv:1509.06928*.

[6] Shivaprasad, S., & Sadanandam, M. (2020). Identification of regional dialects of Telugu language using text independent speech processing models. *International Journal of Speech Technology*, *23*(2), 251-258.

[7] Basu, J., Khan, S., Roy, R., Basu, T. K., & Majumder, S. (2021). Multilingual speech corpus in low-resource eastern and northeastern Indian languages for speaker and language identification. *Circuits, Systems, and Signal Processing*, *40*(10), 4986-5013.

[8] Shah, S. M., Memon, M. U. H. A. M. M. A. D., & Salam, M. H. U. (2020). Speaker recognition for pashto speakers based on isolated digits recognition using accent and dialect approach. *J Eng Sci Technol*, *15*(4), 2190-207.

[9] Das, A. K., Al Asif, A., Paul, A., & Hossain, M. N. (2021). Bangla hate speech detection on social media using attention-based recurrent neural network. *Journal of Intelligent Systems*, *30*(1), 578-591.

[10] Das, A., Kumar, K., & Wu, J. (2021, June). Multi-dialect speech recognition in english using attention on ensemble of experts. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6244-6248). IEEE.

[11] Garain, A., Singh, P. K., & Sarkar, R. (2021). FuzzyGCP: A deep learning architecture for automatic spoken language identification from speech signals. *Expert Systems with Applications*, *168*, 114416.

[12] Hossain, M. F., Hasan, M. M., Ali, H., Sarker, M. R. K. R., & Hassan, M. T. (2020, December). A machine learning approach to recognize speakers region of the united kingdom from continuous speech based on accent classification. In *2020 11th International Conference on Electrical and Computer Engineering (ICECE)* (pp. 210-213). IEEE.

[13] Kaur, J., Singh, A., & Kadyan, V. (2021). Automatic speech recognition system for tonal languages: State-of-the-art survey. *Archives of Computational Methods in Engineering*, *28*, 1039-1068.

[13] Krishna, G. R., Krishnan, R., & Mittal, V. K. (2020, December). Foreign accent recognition with South Indian spoken english. In *2020 IEEE 17th India Council International Conference (INDICON)* (pp. 1-5). IEEE.

[15] Londhe, N. D., & Kshirsagar, G. B. (2018). Chhattisgarhi speech corpus for research and development in automatic speech recognition. *International Journal of Speech Technology*, *21*, 193-210.

[16] Nallasamy, U. (2016). *Adaptation techniques to improve ASR performance on accented speakers* (Doctoral dissertation, Carnegie Mellon University).

[17] Wagner, H. (2019). Austrian Dialect Classification Using Machine Learning.

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title:

Regional Accent Classification in Bangladesh: A Supervised Machine Learning Approach

Student ID: 203-15-14491

203-15-14470

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to identify a real-life complex engineering problem for the Final Year Design Project	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for the Final Year Design Project	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish this goal for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
Phase -II		
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public service standards, as well as considering socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze regional languages in different voice factors along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**

CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**
CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**
CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	Our Project requires data collection manually and we have to know the design of a machine learning-based model which covers	**Page no:** 13-14
				After completing our project, we will make a web application	**Page no:** 13

				We have studies several research papers that related with our works which covers	**Page no:** 7-11

2.	**EP2: Range of Conflicting Requirements**	Yes	CO2	For best accuracy we need raw data from regional people this can be challenging	**Page no:** 5-6
3.	**EP3: Depth of analysisrequired**	Yes	CO2	We will proceed to machine learning model creation after we have a firm understanding of the data. The intricacy of the data and the intended result will determine which model type is used	**Page no:** 6-7
4.	**EP4: Familiarity o Issues**	Yes	CO8	The model's performance on entirely unseen regional accents or speaking styles remains to be explored	**Page no:** 11-12
5.	**EP5: Extends of application codes**	yes	CO5	The XGBoost model could be deployed for real-time spoken language identification in applications like customer service	**Page no:** 19,20, 21,22,23, 24,25,26

6.	**EP6: Extends of stakeholders involved and conflicting**	No	CO8	N/A	N/A
7.	**EP7: Interdependence**	Yes	CO5	This project relies on accurate speech data collection and effective machine learning models for achieving high dialect identification.	**Page no:** 14,15, 16,17,18

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	This project utilizes a combination of speech data from Bangladesh, data augmentation techniques, and various machine learning models for regional accent identification.	**Page no:** 26,27,28,29
2.	**EA2: Level of interaction**	No		N/A	N/A

3.	**EA3: Innovation**	No		N/A	N/A
4.	**EA4: Consequences for society and the environment**	Yes		While positive for accessibility, this project's environmental impact is minimal, with societal considerations focusing on dialect preservation and potential biases in identification.	**Page no:** 4-5
5.	**EA-5: Familiarity**	No		N/A	N/A

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project conduct a cost analysis and economic assessment, and use appropriate project management strategies to guarantee the project's financial sustainability and effective completion.	**Page no:** 29

2	CO9	Yes	By putting patient privacy first, getting informed consent, and openly recording the research process, the project demonstrates adherence to ethical principles. It also ensures responsible knowledge dissemination and societal well-being through the moral application of language processing technologies. which comply with **CO9**.	**Page no:** 45
3	CO10	yes	The project team must be capable of operating proficiently both individually and as team members or leaders, effectively collaborating across diverse and interdisciplinary teams.	**Page no:** 26,27,28,29
4	CO12	Yes	The significance of self-directed and continuous learning to stay up to date with developments in language management and machine learning technologies is highlighted by this study.	**Page no:** 5

REGIONAL ACCENT CLASSIFICATION IN BANGLADESH: A SUPERVISED MACHINE LEARNING APPROACH

ORIGINALITY REPORT

8%	5%	1%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Higher Education Commission Pakistan Student Paper	2%
2	Submitted to Daffodil International University Student Paper	1%
3	thesai.org Internet Source	1%
4	rsisinternational.org Internet Source	<1%
5	Submitted to BITS, Pilani-Dubai Student Paper	<1%
6	0-www-mdpi-com.brum.beds.ac.uk Internet Source	<1%
7	Submitted to General Sir John Kotelawala Defence University Student Paper	<1%
8	www.mdpi.com Internet Source	<1%