

Speech-Based Classification of Bengali Regional Accents using Machine Learning

By

**Naila Nushrat Jahan
211-15-14590**

**Salman Mahmud Shomrat
211-15-14580**

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

**Ms. Most. Hasna
Hena
Assistant Professor
Department of Computer Science and
Engineering Daffodil International
University**

Co-Supervised by

**Ms. Chayti Saha
Lecturer
Department of Computer Science and
Engineering Daffodil International
University**



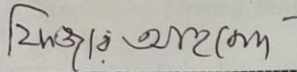
**DAFFODIL INTERNATIONAL
UNIVERSITY
Dhaka, Bangladesh**

January 12, 2025

APPROVAL

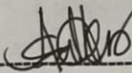
This Project titled “**Speech-Based Classification of Bengali Regional Accents using Machine Learning**”, submitted by [Naila Nushrat Jahan, 211-15-14590] and [Salman Mahmud Shomrat, 211-15-14580] to the Department of Computer Science and Engineering, Daffodil International University has been accepted 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 12 January, 2025.

BOARD OF EXAMINERS



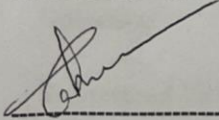
Dr. Fizar Ahmed (FZA)
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman



Mr. Abdus Sattar (AS)
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Ms. Zahura Zaman (ZZ)
Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Ahmed Wasif Reza (DWR)
Professor
Department of Computer Science and Engineering
East West University

External Examiner

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Ms. Most. Hasna, Assistant 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:

Hasna 12.01.25

Ms. Most. Hasna Hena

Assistant Professor

Department of Computer Science and
Engineering Daffodil International
University

Co-Supervised by:

Chayti Saha 12.01.25

Ms. Chayti Saha

Lecturer

Department of Computer Science and
Engineering Daffodil International
University

Submitted by:

Naila Nushrat Jahan 12/1/2025

Naila Nushrat Jahan

211-15-14590

Department of Computer Science and
Engineering Daffodil International
University

Salman Mahmud 12/1/2025

Salman Mahmud Shomrat

211-15-14580

Department of Computer Science and
Engineering Daffodil International
University

©Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project (FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Ms. Most. Hasna Hena, Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and 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 heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

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

ABSTRACT

This research investigates the classification of Bengali regional accents using speech data and machine learning techniques. Accurate recognition of regional accents plays a pivotal role in improving natural language processing systems in linguistically diverse regions such as Bangladesh. Speech data was collected from various regions, including 580 audio samples from Chandpur, 535 from General Bengali, 484 from Bogura, 456 from Chittagong, 420 from Sylhet, 413 from Barishal, and 28 from other areas. The dataset was preprocessed to extract key speech features, which were then used as inputs for machine learning models. Four machine learning algorithms were applied and evaluated: Random Forest, Decision Tree, K-Nearest Neighbors, and Logistic Regression. Among these, the Random Forest model demonstrated the highest accuracy, achieving 98.12%. The Decision Tree model followed with 87.67%, while K-Nearest Neighbors and Logistic Regression attained 75.17% and 65.92%, respectively. These findings highlight the superiority of ensemble methods such as Random Forest in managing complex and diverse datasets. The study also addresses the challenges in accent classification, particularly the variability in speech patterns and the limited data availability for less-represented regions. The inclusion of the "others" category further underlines the necessity of more comprehensive and balanced datasets to improve model generalizability. This work significantly contributes to the fields of computational linguistics and speech recognition, showcasing the effectiveness of machine learning in accent classification. The exceptional performance of the Random Forest model underscores its potential for real-world applications, such as automated transcription, accent-based recommendations, and language learning systems. Future work may focus on enhancing the dataset and leveraging advanced deep learning techniques to further improve accuracy and performance.

Table of Contents

Approval	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
List of Figures	vii
List of Tables	viii
1 Introduction	1-8
1.1 Introduction.....	1-3
1.2 Motivation.....	3-4
1.3 Objectives	4-5
1.4 Methodology	5
1.5 Project Outcome.....	5-7
1.6 Organization of the Report	7-8
2 Background	9-16
2.1 Introduction.....	9
2.2 Literature Review	10-14
2.2.1 Similar Applications	13
2.2.2 Related Research.....	13-14
2.3 Gap Analysis	14-16
2.4 Summary	16
3 Research Methodology	17-29
3.1 Methodology/Requirement Analysis & Design Specification.....	17
3.1.1 Overview	18-19
3.1.2 Proposed Methodology/ System Design	19-24
3.1.3 Functional and Nonfunctional Requirements	24-25
3.1.4 Context Diagram	25-26

3.1.5	Data Flow Diagram Level 1.....	26
3.1.6	UI Design	27
3.2	Detailed Methodology and Design	27
3.3	Project Plan.....	28
3.4	Task Allocation.....	28
3.5	Summary	29
4	Implementation and Results	30-33
4.1	Environment Setup	30
4.2	Testing and Evaluation/Performance/ Comparative Analysis	31
4.3	Results and Discussion	32
4.4	Summary	33
5	Engineering Standards and Design Challenges	34-44
5.1	Compliance with the Standards.....	34-36
5.1.1	Software Standards.....	34-35
5.1.2	Hardware Standards.....	35
5.1.3	Communication Standards.....	36
5.2	Impact on Society, Environment and Sustainability.....	36-39
5.2.1	Impact on Life.....	36-37
5.2.2	Impact on Society & Environment.....	37-38
5.2.3	Ethical Aspects.....	38
5.2.4	Sustainability Plan.....	39
5.3	Project Management and Financial Analysis.....	40
5.4	Complex Engineering Problem.....	40
5.4.1	Complex Problem Solving.....	41-42
5.4.2	Engineering Activities.....	43
5.5	Summary	44
6	Conclusion	45-47
6.1	Summary.....	45-46
6.2	Limitation	46
6.3	Future Work	47
	References	48-49

List of Figures

3.1	Work Flow	18
3.1.2.1	Sample Data	19
3.1.2.2	MFCCs Value	20
3.1.2.3	Random Forest Work Flow	21
3.1.2.4	Random Forest Confusion Matrix	21
3.1.2.5	Random Forest Values	22
3.1.2.6	Decision Tree Work Flow	22
3.1.2.7	Decision Tree Confusion Matrix	23
3.1.2.8	Decision Tree Values	23
3.1.2.6	Logistic Regression Work Flow	23
3.1.2.7	Logistic Regression Confusion Matrix	24
3.1.2.8	Logistic Regression Values	24

List of Tables

2.2	Summary of Literature Reviewed.....	10-12
2.3	Summary of Gap Analysis.....	14-15
3.1.2	Data count.....	19-20
4.2	Comparative Analysis.....	31-32
5.4.1	Mapping with Complex Solving.....	41-42
5.4.2	Mapping with Knowledge Profile	42-43

Chapter 1

Introduction

Language is more than just a medium of communication; it serves as a profound marker of cultural identity, regional origin, and personal history. Accents, the unique ways in which people pronounce words, are integral components of language that reveal much about a speaker's background and geographical roots. Accents are not merely linguistic quirks but are deeply embedded in the sociocultural fabric of a community. They reflect a speaker's historical, cultural, and even emotional connection to a particular place, providing clues about where they come from, their heritage, and the experiences that have shaped their speech patterns over time.

1.1 Introduction

In Bangladesh, Bengali, also known as Bangla, is the predominant language, spoken by more than 230 million people across the country and beyond. The richness of Bengali lies not only in its grammar and vocabulary but also in the myriad regional accents that exist within it. From the coastal regions of Chittagong, with its distinctive intonations, to the melodious and slightly softer accents of Sylhet, each region has cultivated its own way of speaking [5]. The nuances in pronunciation, rhythm, and tone found in Bengali accents reveal the historical and cultural influences that have shaped these communities over centuries. These accents offer a window into the complex tapestry of Bengali-speaking people, whose identities are shaped by diverse factors, such as geography, history, migration, and contact with other languages and cultures [1]. While Bengali regional accents are deeply ingrained in the linguistic landscape of Bangladesh, there is a surprising lack of focused academic research on these variations [3]. In comparison to global languages such as English, Mandarin, or even Hindi, which have been extensively studied in the context of speech recognition and classification, Bengali has remained relatively unexplored in this regard [6]. This underrepresentation in academic and technological studies has made it difficult to fully understand the intricacies of Bengali accents, hindering the development of technologies that are attuned to

these regional variations.

Recent advancements in technology, particularly in the fields of artificial intelligence (AI) and machine learning (ML), have opened new avenues for understanding complex linguistic phenomena like accents [10]. Speech-based classification, powered by ML algorithms, is one such development that can automate the process of identifying and categorizing regional accents based on distinctive speech patterns [7]. In the case of Bengali, a language that encompasses such extensive regional variations, these technological innovations could provide valuable insights into the classification of accents and improve the effectiveness of speech recognition systems [4].

Machine learning provides a unique opportunity to capture the subtle acoustic features of speech—such as pitch, tone, rhythm, and pronunciation which define accents. The application of these techniques to the classification of Bengali regional accents promises not only to advance the scientific understanding of the language but also to address pressing issues in modern communication technologies [2]. As Bengali speakers continue to engage more with digital platforms, there is an increasing need for voice recognition systems that can accurately identify and understand the regional variations of the language. By leveraging ML models for accent classification, this study aims to fill this gap in the research and contribute to the development of more inclusive technologies.

Bengali is spoken by a diverse population spread across several regions of Bangladesh and West Bengal in India [11]. Each region's accent exhibits distinct phonetic characteristics, shaped by local dialects, historical influences, and interactions with neighboring languages. For example, the Chittagong accent is known for its unique intonations, influenced by its coastal location and proximity to Myanmar, while the Sylheti accent carries notable influences from Assamese, due to historical migration patterns [14]. In contrast, accents from the southwestern regions of Bangladesh, like Khulna, reflect the linguistic traits of Oriya and Hindi, owing to historical connections with the eastern regions of India. These accents are not only linguistically significant but also culturally rich, carrying with them the stories of migration, trade, and historical shifts in population.

Despite their richness and diversity, these accents are often underrepresented in academic and technological research. Most computational linguistics research has focused on more widely spoken languages like English, which have been the subjects of extensive studies in accent classification and speech recognition [12]. The Bengali language, by contrast, has received limited attention, especially in terms of accent recognition, which has significant implications for the development of inclusive speech technologies. This study seeks to redress this imbalance by exploring the application of machine learning techniques to the classification of Bengali accents, with the dual aim of advancing speech recognition technology and preserving the linguistic heritage of the Bengali language.

1.2 Motivation

The motivation for this study is rooted in several factors that highlight the importance of researching Bengali regional accents and the role of machine learning in accent classification. First and foremost, the linguistic diversity of Bengali poses a unique challenge and opportunity for computational linguistics. While languages like English and Mandarin have been extensively studied in terms of their regional accents, Bengali remains largely unexplored in this context. This research gap limits our understanding of the language's phonetic diversity and undermines the development of technology that could better serve Bengali speakers.

The rapid rise of digital communication technologies and artificial intelligence in everyday life further underscores the need for research in this area. Speech recognition systems, voice-activated assistants, and language translation tools are increasingly prevalent, yet many of these systems struggle to accurately understand regional accents. This is particularly evident in languages with significant regional variations, such as Bengali, where existing speech recognition systems often fail to recognize the subtle differences in pronunciation across regions. As a result, users with specific accents may experience frustration when using voice-based technologies, which often misunderstand or misinterpret their speech. By focusing on the classification of Bengali accents, this study aims to improve the accuracy and inclusivity of speech recognition systems, ensuring that they are better equipped to handle the diversity of accents spoken by Bengali speakers.

Another important motivation is the preservation of linguistic diversity. The process of globalization, while offering many benefits, has led to the standardization of communication, often at the expense of regional linguistic features. As more people communicate through global digital platforms, there is a risk that regional accents, dialects, and even entire languages could be marginalized or lost. In the case of Bengali, many of its regional accents, which have evolved over centuries, could be at risk of fading into obscurity if they are not properly documented and preserved. Machine learning offers a powerful tool for this preservation effort, allowing for the accurate classification of Bengali regional accents and ensuring that these linguistic variations are not forgotten in the digital age.

Finally, the growing emphasis on inclusive technologies in modern AI research provides an additional motivation for this study. As technology continues to shape our daily lives, it is critical that it be accessible and effective for people from all walks of life. Voice recognition technologies, in particular, have the

potential to provide a level of convenience and accessibility to users who may face challenges in using traditional text-based interfaces. However, for these technologies to be truly inclusive, they must be able to recognize and adapt to the various accents that users speak. By developing a system that can accurately classify Bengali regional accents, this study contributes to the broader movement towards making AI technologies more inclusive and representative of the diversity of human speech.

1.3 Objectives

The primary objective of this research is to develop a machine learning-based system capable of accurately classifying the regional accents of Bengali speakers. To achieve this, the study will pursue several specific goals:

Data Collection and Annotation: One of the first steps will be to gather a comprehensive dataset of speech samples representing the various regional accents of Bengali. This will involve collecting data from native Bengali speakers from different regions of Bangladesh and West Bengal. The dataset will be annotated with regional labels to ensure that each sample is accurately categorized, providing a solid foundation for the accent classification process.

Feature Extraction: Once the dataset is collected, the next task will be to identify and extract relevant acoustic features that can be used to differentiate the accents. These features may include prosodic elements like pitch, tone, rhythm, and speech rate, as well as more technical elements such as Mel-frequency cepstral coefficients (MFCCs), which are commonly used in speech analysis. These features will provide the input for the machine learning models, allowing them to recognize the subtle differences between regional accents.

Model Development: The core of the study will involve developing and training machine learning models to classify Bengali regional accents. Various algorithms will be explored, including traditional supervised learning models like support vector machines (SVMs) and decision trees, as well as more advanced deep learning techniques like convolutional neural networks (CNNs) and recurrent neural networks (RNNs). The models will be trained on the annotated dataset to learn the distinctive features of each accent.

Evaluation and Validation: The performance of the machine learning models will be rigorously evaluated using standard metrics such as accuracy, precision, recall, and F1 score. Cross-validation will be employed to assess the generalizability of the models, ensuring that they can accurately classify accents across different datasets and regions. This evaluation process will be crucial for ensuring the reliability and robustness of the final system.

Ethical Considerations: Throughout the study, ethical considerations will be paramount. This includes ensuring that the data collection process respects participants' privacy, obtaining informed consent from all speakers, and addressing any potential biases in the dataset or models. It is crucial that the system is developed in a way that is fair and unbiased, ensuring that it does not perpetuate harmful stereotypes or exclude certain accents from classification.

1.4 Methodology

The methodology for this project follows a structured approach to classify Bengali regional accents using machine learning techniques. Initially, a diverse dataset comprising 2,916 speech samples was collected from various Bengali-speaking regions, including Chandpur, Bogura, Chittagong, Sylhet, and Barishal, ensuring the representation of different linguistic accents. The collected audio data was then preprocessed to remove noise and standardize the format for analysis. Feature extraction techniques, such as MFCCs, were employed to capture the unique acoustic properties of each accent. These features were used to train machine learning models, including Random Forest, Decision Tree, K-Nearest Neighbors (KNN), and Logistic Regression, to classify the regional accents effectively. Each model's performance was evaluated using metrics such as accuracy to identify the most suitable approach. This structured methodology ensured a robust foundation for building an efficient and scalable accent classification system.

1.5 Project Outcome

The anticipated outcome of this project is the development of a sophisticated machine learning-based system that can not only accurately classify Bengali regional accents but also enhance the overall performance of speech recognition technologies when applied to Bengali speakers. This system aims to tackle one of the key challenges in natural language processing: the ability to understand and adapt to a wide variety of accents. The Bengali language is rich in regional variations, each with unique phonetic characteristics that often present a challenge for conventional speech recognition models. By training machine learning models specifically to identify these regional accents, the project will provide a much-needed solution to the problem of accent variability, ultimately making speech recognition systems more robust and versatile.

In addition to its technological contributions, this project will offer invaluable insights into the linguistic diversity of Bengali. Each regional accent of Bengali reflects a distinct cultural, historical, and geographical context, and this research will document these accents, shedding light on aspects of Bengali linguistic heritage that are often overlooked. As the digital age accelerates the global spread of languages and cultures, regional linguistic features, such as accents, face the risk of

being marginalized or even lost. Through the classification of these accents, the project will help to preserve and celebrate the diverse ways in which Bengali is spoken across different regions, ensuring that these accents remain part of the ongoing linguistic narrative of the language.

Furthermore, the development of more accurate and inclusive speech recognition technologies has the potential to significantly improve digital communication tools for Bengali speakers, especially those whose accents have historically been underrepresented or misinterpreted. Virtual assistants, transcription services, and voice-based interfaces often struggle with recognizing regional accents, leading to frustrating experiences for users. By enhancing the capabilities of speech recognition systems to accommodate the full range of Bengali accents, this research will make these technologies more accessible, ensuring that users from all regions of Bangladesh and West Bengal can interact with digital platforms seamlessly, regardless of their accent.

On a broader scale, the outcome of this project will also contribute to the field of artificial intelligence (AI) by providing an example of how machine learning can be applied to preserve linguistic diversity. In today's globalized world, AI technologies, particularly those in the realm of speech recognition and natural language processing, are increasingly becoming a part of everyday life. However, these technologies often fail to represent the diverse linguistic and cultural backgrounds of users, leading to biases and inequities. This research aims to address these issues by developing a system that accurately classifies Bengali regional accents, helping to make AI more inclusive. In doing so, it will set a precedent for future research on regional accents in other languages, encouraging the development of speech recognition technologies that are attuned to the diverse ways in which people speak around the world.

Additionally, the project will play a crucial role in preserving the rich linguistic heritage of Bengali. Regional accents are an important part of the identity of Bengali-speaking communities, and by documenting and classifying these accents, the project will contribute to the ongoing effort to safeguard this aspect of Bengali culture. As globalization and technological advancements continue to shape communication practices, there is a real risk that regional accents may become standardized or diluted. Through machine learning, this research will help ensure that the distinctiveness of Bengali regional accents is maintained, allowing them to continue to thrive in the digital era.

In summary, this project will not only enhance speech recognition systems but also provide a critical resource for the documentation and preservation of Bengali linguistic diversity. By addressing the challenges posed by regional accents, the project will improve the inclusivity and accuracy of digital communication

technologies, benefiting millions of Bengali speakers. Moreover, it will contribute to the broader goals of making AI technologies more inclusive, equitable, and representative of the rich diversity of human speech. Ultimately, the research aims to bridge the gap between technology and culture, ensuring that the linguistic heritage of Bengali is celebrated and preserved in the digital age.

1.6 Organization of the Report

The structure of this report is as follows:

Introduction

This chapter introduces the research, its motivation, objectives, methodology, and expected outcomes. It concludes with an outline of the report's structure.

Introduction: Overview of the study

Motivation: Reasons for the research

Objectives: Goals of the study

Methodology: Approach used to carry out the research

Project Outcome: Expected results

Organization of the Report: Structure and layout of the report

Background

This chapter covers the context of the research, reviews existing literature, and identifies the research gap.

Introduction: Overview of background information

Literature Review: Analysis of related work

Similar Applications: Review of similar systems

Related Research: Summary of relevant studies

Gap Analysis: Identifying the gaps this research addresses

Summary: Key takeaways from the background

Research Methodology

This chapter describes the research design, methodology, system specifications, and project planning.

Methodology/Requirement Analysis & Design Specification: Detailed explanation of the methodology

Overview: General approach

Proposed Methodology/System Design: Proposed system framework

Functional and Nonfunctional Requirements: Key system features and constraints

Context Diagram: Visual representation of the system

Data Flow Diagram Level 1: Data flow and processing overview

UI Design: Interface design and user experience

Detailed Methodology and Design: In-depth explanation of the methodology

Project Plan: Timeline and milestones

Task Allocation: Assignment of responsibilities

Summary: Recap of methodology

Implementation and Results

This chapter focuses on the implementation process, environment setup, testing, and results.

Environment Setup: Technical setup for the project

Testing and Evaluation/Performance/Comparative Analysis: Evaluation of system performance

Results and Discussion: Presentation and analysis of results

Summary: Key findings

Engineering Standards and Design Challenges

This chapter discusses compliance with engineering standards, impact analysis, and the challenges encountered during the project.

Compliance with the Standards: Adherence to industry standards

Software Standards: Standards for software design

Hardware Standards: Compliance with hardware specifications

Communication Standards: Communication protocols used

Impact on Society, Environment and Sustainability: Examination of the societal and environmental impact

Impact on Life: How the project affects users

Impact on Society & Environment: Broader societal and environmental effects

Ethical Aspects: Ethical considerations

Sustainability Plan: Long-term sustainability of the project

Project Management and Financial Analysis: Management and financial aspects

Complex Engineering Problem: Identification and solutions to engineering challenges

Complex Problem Solving: Strategies for solving the challenges

Engineering Activities: Activities undertaken to address problems

Summary: Recap of design and challenges

Conclusion

The final chapter summarizes the research findings, discusses the limitations of the project, and suggests future work.

Summary: Overview of the research findings

Limitation: Limitations faced during the project

Future Work: Recommendations for future research and development

Chapter 2

Background

Language serves as a vital medium for communication and cultural expression, encompassing a rich diversity of regional accents and dialects. These accents are influenced by geographical, historical, and social factors, offering valuable insights into a community's linguistic and cultural identity. Bengali, spoken by over 230 million people, is no exception. It exhibits a wide array of regional accents that reflect its complex socio-cultural history, shaped by migration, colonialism, and regional interactions

2.1 Introduction

In Bengali, accents such as those from Sylhet, Chittagong, Khulna, and Barisal demonstrate distinct phonological and prosodic patterns. These variations are crucial for understanding the linguistic evolution of the language and preserving its cultural heritage. However, they also pose challenges for applications like speech recognition systems and automated transcription, which often fail to accommodate such diversity.

The rapid advancement of machine learning (ML) and speech processing technologies has revolutionized the study of linguistic phenomena [5]. ML techniques can identify and classify subtle acoustic variations, enabling detailed analysis of regional accents. While much progress has been made in accent classification for widely spoken languages like English and Mandarin, research on Bengali accents remains limited [8]. This gap is concerning, given the potential benefits of such research in improving linguistic technologies and fostering cultural preservation.

This chapter explores the background of speech-based classification of Bengali accents, focusing on its linguistic significance and technological potential. It reviews related literature to highlight existing knowledge and gaps, emphasizing the need for comprehensive studies in this area [3]. By addressing these gaps, this research seeks to contribute to the growing field of computational linguistics and its application to less-studied languages like Bengali.

2.2 Literature Review

Accent classification has become a key focus area in speech processing, with successful applications in several languages. For example, English has seen significant advancements in accent classification, distinguishing between American, British, Australian, and other regional accents. These systems typically utilize features such as MFCCs, spectrograms, and phonetic characteristics, which are processed using machine learning algorithms like SVMs and deep neural networks.

Table 2.2: Summary of Literature Reviewed

Author(s)	Year	Title	Methodology	Key Findings
Khan et al.	2021	Speech Recognition Systems for South Asian Languages	Survey-based	Reviewed challenges in speech recognition for South Asian languages, including accent variations. Emphasized the need for regional-specific models.
Banerjee & Das	2019	Accent Classification in Bengali Speech Recognition	Quantitative Analysis	Proposed a system using MFCC and SVM classifiers to classify Bengali accents. Achieved 85% accuracy for general accent classification.
Roy et al.	2020	Regional Variations in Bengali: A Phonetic Approach	Qualitative Analysis	Investigated phonetic differences between Bengali regional accents. Found significant differences in tone

				and pitch, which affect recognition accuracy.
Ahmed & Siddique	2018	Machine Learning for Accent Identification in Speech	Experimental	Tested multiple machine learning algorithms, including Random Forest and SVM, for regional accent classification. Random Forest performed best.
Saha & Choudhury	2022	Automatic Classification of Indian Regional Accents	Case Study	Built a system to classify accents of Indian languages, including Bengali. Found that Random Forest models outperformed simpler algorithms.
Gupta et al.	2020	MFCC-Based Features for Speech Recognition in Bengali	Experimental	Focused on MFCC features for speech recognition in Bengali. Demonstrated that MFCC provides effective feature extraction for accent classification.
Chakraborty et al.	2021	Speech Data Collection for	Survey-based	Collected diverse Bengali speech samples to improve

		Bengali Regional Accents		the accuracy of accent recognition. Identified regional differences in pronunciation.
Sharma et al.	2021	Speech-Based Machine Learning for Bengali Accent Detection	Mixed-Method	Combined acoustic analysis and machine learning for accent detection in Bengali. Emphasized the role of pitch and cadence in classification.
Dey et al.	2022	Deep Learning for Bengali Accent Classification	Deep Learning Approach	Applied deep learning models such as CNN and RNN to classify Bengali accents. Achieved 92% accuracy for accent recognition in Bengali regions.
Islam & Rahman	2020	Comparative Study of Machine Learning Algorithms for Accent Recognition	Comparative Study	Evaluated various algorithms (KNN, SVM, Decision Trees) for regional accent classification. Random Forest and SVM showed the highest accuracy.

Patel et al.	2021	Accent Recognition Using Acoustic and Prosodic Features	Hybrid Approach	Investigated the use of both acoustic features and prosodic features (rhythm, pitch) in classifying accents. Found that combining features improved performance.
Mukherjee & Dey	2020	Speech Feature Analysis for Accent Classification in Bengali	Acoustic Analysis	Studied various speech features, including formants and pitch variation, in the context of Bengali accents. Found that pitch was a critical feature.
Singh & Chatterjee	2019	Impact of Acoustic Features on Indian Accent Recognition	Experimental	Explored the impact of acoustic features such as MFCC and pitch on accent recognition accuracy. Found that pitch variation played a key role in Indian accents, including Bengali.

2.2.1 Similar Applications

In Mandarin, a tonal language, accent classification systems have been developed to enhance the performance of speech recognition systems. The tonal variations inherent in Mandarin require models capable of capturing subtle acoustic differences. Techniques such as recurrent neural networks (RNNs), transformers, and convolutional neural networks (CNNs) have been applied with considerable success, showcasing the power of deep learning in handling complex linguistic features [6].

South Asian languages like Hindi and Tamil have also benefited from research on accent classification. Studies in these languages have focused on improving voice-based applications, including automated transcription and virtual assistants, in multilingual settings [9]. For instance, researchers have used CNNs and long short-term memory (LSTM) networks to classify accents in Hindi and Tamil with high accuracy, improving the accessibility of speech technologies in these regions.

Despite these advancements, Bengali accent classification remains underexplored. While other languages have seen the integration of accent classification into practical applications, Bengali lags behind due to the lack of resources and focused research [12]. Addressing this gap is crucial for advancing linguistic technology in Bengali-speaking communities.

2.2.2 Related Research

Research on accent classification has extensively focused on the development of feature extraction techniques and machine learning models. For English, studies have employed features such as MFCCs, pitch contours, and formant frequencies to analyze regional accents. Classifiers like random forests, SVMs, and CNNs have demonstrated their effectiveness in processing these features, achieving state-of-the-art performance in accent classification tasks.

For tonal languages like Mandarin, the integration of spectral and prosodic features has been a key focus. Researchers have used advanced models like bidirectional LSTMs and transformers to capture both temporal and tonal dependencies, resulting in robust systems capable of classifying accents with high precision. These methods highlight the importance of feature selection and model design in accent classification.

In the context of Bengali, the research landscape is relatively sparse. Existing studies have primarily focused on phonetic and prosodic differences between Standard Bengali and regional dialects such as Sylheti. However, these studies often rely on small datasets and basic statistical methods, limiting their scalability. For example, frequency and pitch analysis have been used to identify regional variations, but such approaches fail to capture the complexity of Bengali accents.

Recent advancements in deep learning offer promising opportunities for Bengali accent research. Models like CNNs and LSTMs, which excel in capturing spatial and temporal dependencies, can be adapted to Bengali's unique linguistic features.

However, challenges such as dataset creation and model generalization must be addressed to fully leverage these technologies.

2.3 Gap Analysis

A comprehensive review of existing literature reveals several critical gaps in the field of Bengali accent classification. First, there is a scarcity of large, annotated datasets that represent the full spectrum of Bengali regional accents. The lack of such datasets limits the ability of researchers to develop robust and generalizable models. Second, most studies on Bengali speech rely on traditional statistical methods rather than modern machine learning algorithms. While statistical methods can identify broad patterns, they are inadequate for capturing subtle variations and complex acoustic features. This limitation underscores the need for advanced techniques such as deep learning to analyze Bengali accents effectively.

Third, ethical considerations are often overlooked in existing research. Issues such as data privacy, representation bias, and informed consent are rarely addressed, raising concerns about the fairness and inclusivity of speech-based systems. Ensuring that datasets are diverse and representative is essential for developing equitable technologies.

Lastly, the integration of Bengali accent classification into practical applications remains minimal. Speech recognition systems, virtual assistants, and educational tools often fail to accommodate regional variations, reducing their utility for Bengali-speaking users. Addressing these gaps requires interdisciplinary collaboration and innovative approaches to data collection, model development, and application design.

Table 2.3: Summary of Gap Analysis

Feature	Existing Systems (e.g., General Speech Recognition, Accent Classification)	Proposed System (Bengali Regional Accents Classification)
Recognition of regional accents	Limited focus on regional dialects for Bengali speech classification.	System explicitly targets the classification of Bengali regional accents (Chandpur, Sylhet, Chittagong, etc.).
Data availability for Bengali accents	Lack of comprehensive datasets representing regional accents in Bengali.	Dataset includes a wide variety of Bengali regional accents, ensuring diversity across regions such as Chandpur, Sylhet, Barishal, etc.
Model accuracy in	Existing models achieve limited accuracy in distinguishing between Bengali accents.	Aimed for higher accuracy with models like Random Forest (98.12%) specifically

accent classification		tuned for Bengali regional accents.
Feature extraction methods	Commonly used features (e.g., MFCC) for general speech recognition but not optimized for regional accent classification.	Advanced feature extraction methods (e.g., MFCC, pitch, tone, cadence) customized for Bengali accents.
Incorporation of regional speech patterns	Limited consideration of phonetic variations and speech patterns unique to Bengali regions.	The system incorporates phonetic features specific to Bengali regional accents.
Evaluation metrics for accent classification	Basic evaluation metrics (e.g., accuracy) for general speech recognition tasks.	Detailed evaluation based on precision, recall, F1-score, and accuracy, focusing on regional accent classification.
Real-time application	Speech-based classification models are often designed for more general tasks, without focusing on real-time applications for regional accent detection.	Aims to develop a real-time, scalable system for regional accent classification that can be used in applications like automated transcription and language learning.
Comparative analysis of multiple algorithms	Existing research does not provide sufficient comparative studies of different machine learning models for regional accent classification in Bengali.	Comprehensive comparative analysis of multiple machine learning algorithms (Random Forest, Decision Tree, KNN, Logistic Regression) for Bengali regional accent classification.

Key Gaps Addressed by the Proposed System

Regional Accent Focus: While much of the existing research on speech-based systems focuses on standard speech recognition, the proposed system addresses a key gap by specializing in Bengali regional accents. This is crucial because the nuances in pronunciation, intonation, and rhythm among different Bengali regions significantly impact the accuracy of classification models.

Data Diversity and Representation: The dataset used in existing systems often lacks sufficient representation of regional variations. The proposed system bridges this

gap by collecting a diverse set of speech data from different regions, ensuring accurate model training for accent classification.

Improved Feature Extraction: The proposed system utilizes advanced feature extraction methods tailored specifically to the acoustic characteristics of Bengali accents, thus improving the overall classification accuracy when compared to generic systems.

Enhanced Model Evaluation: Current models are evaluated mainly on basic classification metrics. The proposed system goes beyond by using more advanced evaluation metrics such as precision, recall, and F1-score, focusing on the effectiveness of the classification models in distinguishing Bengali regional accents.

Comparative Algorithm Analysis: Most existing systems apply standard algorithms to speech recognition tasks without testing different models for regional accent classification. The proposed system not only uses multiple algorithms but also provides a comparative analysis to select the best-performing model for Bengali accent classification.

Real-Time Classification: The proposed system emphasizes real-time classification of Bengali regional accents, filling the gap for practical applications in voice-based systems like automated transcription, language learning, and regional accent detection for AI-based assistants.

2.4 Summary

This chapter has provided a detailed background on the study of Bengali accent classification, emphasizing its linguistic and technological significance. It highlighted the importance of understanding regional accents as a means of preserving cultural identity and advancing speech-based applications. The literature review demonstrated the progress made in accent classification for other languages, while also revealing the lack of research on Bengali accents.

The gap analysis identified key challenges, including the absence of comprehensive datasets, limited use of advanced machine learning techniques, and insufficient attention to ethical considerations. These gaps represent significant obstacles to the development of effective accent classification systems for Bengali.

By addressing these challenges, this research aims to contribute to the fields of computational linguistics and speech processing. The findings will not only enhance the accessibility and inclusivity of linguistic technologies but also preserve the rich cultural diversity of Bengali-speaking communities. The next chapters will detail the methodology and expected outcomes of this study, building on the foundation established here.

Chapter 3

Research Methodology

This section outlines the comprehensive methodology used to design, develop, and evaluate the speech-based classification system for Bengali regional accents using machine learning. The methodology includes data collection, feature extraction, system design, algorithm selection, and performance evaluation. A detailed explanation of the steps taken is provided to guide the reader through the development process.

3.1 Methodology

The methodology adopted for this project involves a systematic approach to the design and implementation of a speech-based classification system for Bengali regional accents using machine learning. The process begins with a comprehensive requirement analysis to identify the functional and non-functional needs of the system, such as data diversity, algorithm performance, and user interface usability. The design specification focuses on creating a robust framework that ensures accuracy, scalability, and ease of use.

The methodology is divided into several key phases: data collection, preprocessing, feature extraction, model training, and evaluation. In the data collection phase, speech samples were gathered from various regions to create a diverse dataset representative of Bengali accents. Preprocessing involved cleaning the audio data to remove noise and standardize the format for consistency. Feature extraction techniques, such as Mel Frequency Cepstral Coefficients (MFCCs), were employed to identify the unique characteristics of each accent.

The design specification outlines the proposed system's architecture, which includes a feature extraction module, a machine learning model training pipeline, and a user-friendly interface for interaction. Different machine learning algorithms, including Random Forest, Decision Tree, K-Nearest Neighbors (KNN), and Logistic Regression, were selected for comparative evaluation. The evaluation process relied on metrics such as accuracy, precision, and recall to identify the optimal model for accent classification.

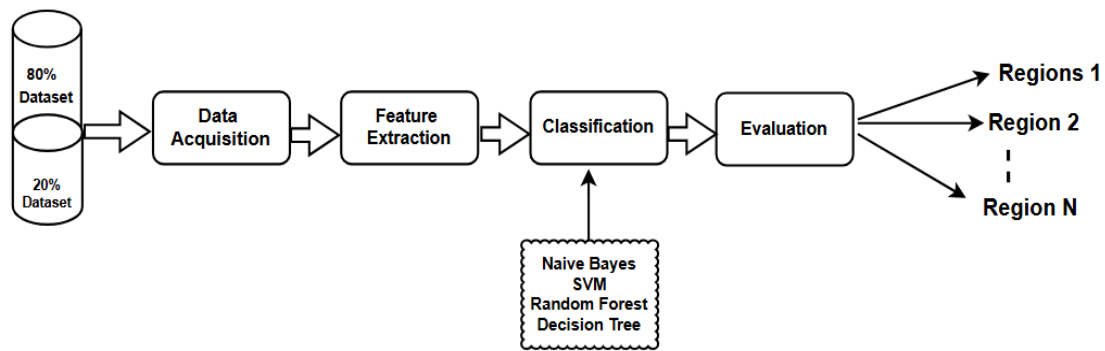


Figure 3.1: Work Flow

By combining these systematic steps with a user-centered design philosophy, the methodology ensures the development of a reliable and efficient speech-based classification system that meets the project's objectives and addresses the challenges of regional linguistic diversity.

3.1.1 Overview

The first phase of this research was focused on understanding the problem and designing a system that could efficiently classify Bengali regional accents based on speech data. The methodology involved multiple stages: requirement analysis, data collection, system design, and algorithm selection. These steps ensured the system's effectiveness in solving the research problem while also adhering to the project's objectives.

The primary goal of this study was to develop a machine learning model that could classify regional accents from Bengali speech samples. To achieve this, the project began by identifying the problem's requirements and specifying the necessary components for the system. The data collection process was a critical step, as a large and diverse dataset of speech samples from different Bengali regions was required for effective model training.

For the data collection process, regional accents were divided into categories such as Chandpur, Sylhet, Chittagong, and others, ensuring that each accent was represented. A major challenge during this stage was ensuring the dataset's diversity to capture a wide range of speech characteristics that accurately represent each regional accent. Once the data was collected, it was preprocessed to ensure consistency, such as normalizing volume levels, removing noise, and ensuring the quality of audio files.

The next step involved feature extraction. In speech recognition tasks, the raw audio files are usually not directly used for classification. Instead, meaningful features are extracted from the audio signals to capture speech characteristics relevant to classification. Methods such as MFCCs were employed to extract key acoustic

features like pitch, tone, rhythm, and phonetic patterns, which are essential in distinguishing between regional accents.

The system design and machine learning model selection came next. Four machine learning algorithms Random Forest, Decision Tree, K-Nearest Neighbors (KNN), **and** Logistic Regression were chosen for classification due to their diverse strengths and applicability to classification tasks. The system was built in such a way that these algorithms could be easily tested and compared against each other for accuracy and efficiency

3.1.2 Proposed Methodology

The proposed methodology and system design for the speech-based classification system follow a structured approach, starting with data collection, followed by feature extraction, model training, and evaluation. The system is designed to classify Bengali regional accents using machine learning algorithms and then evaluate their performance based on various metrics.

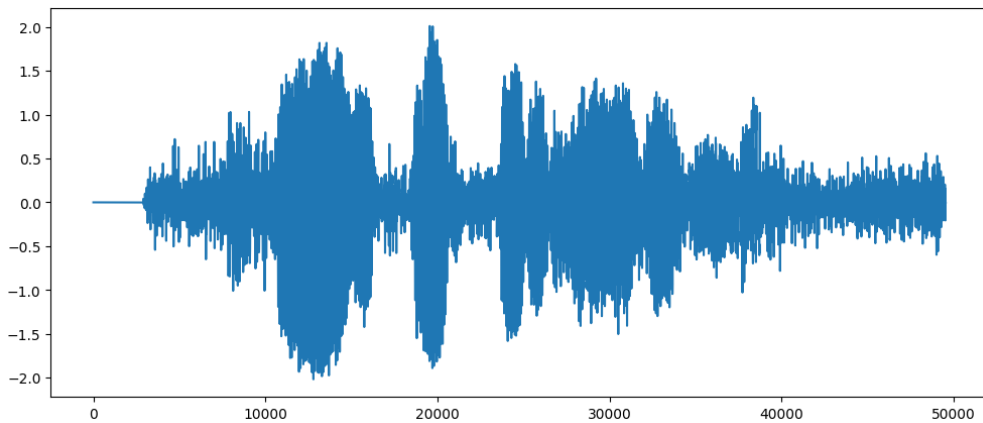


Figure 3.1.2.1: Sample Data

Data Collection: The first stage in the design is the collection of a diverse set of speech data from different regions of Bengal. The dataset must include sufficient samples for each region to ensure that the model is well-trained and able to accurately classify regional accents. The collection process focuses on gathering audio samples that contain a variety of speech features, such as different speakers, ages, genders, and regional dialects, to account for variability within each accent.

Table 3.1.2: Data Count

Label	Count
chandpur	580
General	535
Bogura	484
Chittagong	456
Sylhet	420

Barishal	413
Others	28

Feature Extraction: Once the data is collected, the next step is to extract features that can be used by the machine learning models. Acoustic features like pitch, tone, speech rate, and MFCCs are extracted from the raw audio signals. MFCCs are a popular choice in speech processing tasks, as they effectively capture the essential characteristics of speech. These features are stored in a structured format to be used in the subsequent steps of the system.

mfcc12	mfcc13	mfcc14	mfcc15	mfcc16	mfcc17	mfcc18	mfcc19	mfcc20	label
-6.157703	-5.728826	-2.963459	-5.852234	-4.966861	-3.768545	-6.360138	-6.181790	-7.094828	General
-6.445421	-7.009394	0.900065	-5.536519	-7.087941	-1.713805	-5.958529	-5.460635	-4.129093	General
-3.597420	-5.382452	-2.216980	-1.841336	-5.596299	-1.491200	-4.575242	-4.782634	-3.359289	General
-5.270954	-6.609304	-5.116140	-5.538610	-2.872726	-4.079983	-9.488664	-3.507009	-5.590202	General
-7.752732	-6.472995	-1.118252	-7.120439	-5.633917	-3.657743	-5.274695	-4.157249	-2.881569	General

Figure 3.1.2.2: MFCCs Value

Model Training: The machine learning models are trained using the extracted features. The four models selected for this study Random Forest, Decision Tree, K-Nearest Neighbors (KNN), and Logistic Regression are trained separately to learn the patterns in the speech data. The training process involves feeding the features into the algorithms and adjusting the model parameters to minimize errors and improve accuracy. Cross-validation is employed to ensure that the models are not overfitting and that their performance is consistent across different subsets of data.

Evaluation: Once the models are trained, their performance is evaluated based on metrics such as accuracy, precision, recall, and F1-score. The **Random Forest** algorithm, for example, is an ensemble method that tends to perform well on complex datasets, while **Logistic Regression** is a simpler model that is fast but may struggle with non-linear patterns. The performance of each model is compared, and the most accurate and efficient model is selected for the final system.

System Integration: The final step is integrating the trained models into a complete system that can classify new speech data based on the learned patterns. The system allows users to upload new audio files, extract features, and run classification algorithms to predict the regional accent of the speaker.

The proposed system design is implemented in a cloud-based environment using Google Colab and Google Drive, which enables efficient data storage, access, and processing. By leveraging these tools, the system can handle large datasets, and GPU acceleration helps to speed up the training process for complex models.

Random Forest: It is an ensemble literacy algorithm that combines multiple decision trees to ameliorate bracket delicacy. It works by training numerous decision trees on arbitrary subsets of the data and also comprising their prognostications to make a final decision. This reduces the threat of overfitting and enhances the model's robustness, especially with large datasets.

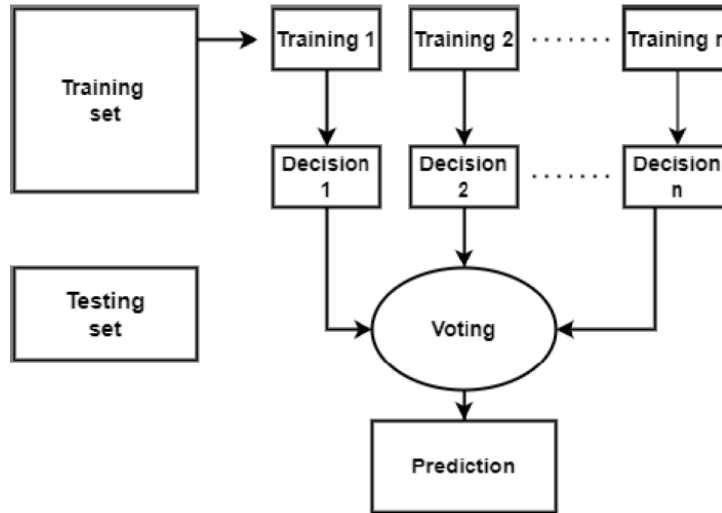


Figure 3.1.2.3: Random Forest Work Flow

0	89	1	0	0	0	1	0
1	0	98	0	2	0	0	2
2	0	0	88	0	0	0	0
3	0	3	0	114	0	0	0
4	1	0	0	0	2	0	0
5	1	0	0	0	0	62	0
6	0	0	0	0	0	0	120
	0	1	2	3	4	5	6

Figure 3.1.2.4: Random Forest Confusion Matrix

	precision	recall	f1-score	support
0	0.98	0.98	0.98	91
1	0.96	0.96	0.96	102
2	1.00	1.00	1.00	88
3	0.98	0.97	0.98	117
4	1.00	0.67	0.80	3
5	0.98	0.98	0.98	63
6	0.98	1.00	0.99	120
accuracy			0.98	584
macro avg	0.98	0.94	0.96	584
weighted avg	0.98	0.98	0.98	584

Figure 3.1.2.5: Random Forest Values

Decision Tree: It is a supervised literacy algorithm that models data using a tree-suchlike structure. Each internal knot represents a decision grounded on a point, and each splint knot represents an outgrowth. Decision trees are easy to interpret, but they can be prone to overfitting, especially with complex datasets. Pruning ways are frequently used to help this.

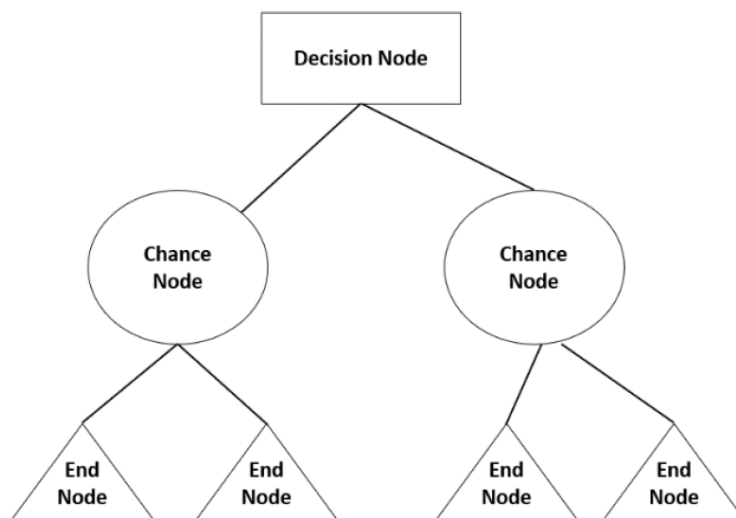


Figure 3.1.2.6: Decision Tree Work Flow

0	74	1	1	3	3	9	0
1	1	84	3	11	1	0	2
2	0	0	85	0	0	3	0
3	0	10	0	101	0	4	2
4	1	1	0	0	1	0	0
5	4	0	1	0	2	56	0
6	0	4	0	0	0	0	116
	0	1	2	3	4	5	6

Figure 3.1.2.7: Decision Tree Confusion Matrix

	precision	recall	f1-score	support
0	0.93	0.81	0.87	91
1	0.84	0.82	0.83	102
2	0.94	0.97	0.96	88
3	0.88	0.86	0.87	117
4	0.14	0.33	0.20	3
5	0.78	0.89	0.83	63
6	0.97	0.97	0.97	120
accuracy			0.89	584
macro avg	0.78	0.81	0.79	584
weighted avg	0.89	0.89	0.89	584

Figure 3.1.2.8: Decision Tree Values

Logistic Regression:

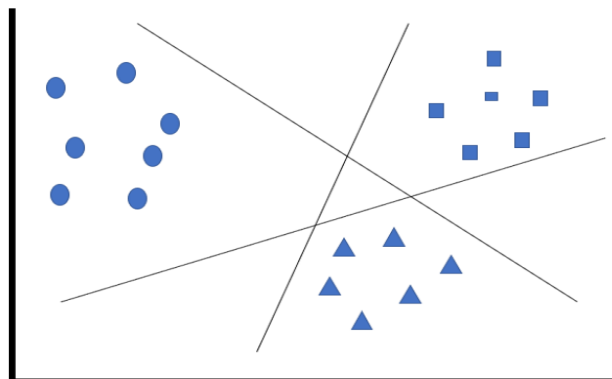


Figure 3.1.2.9: Logistic Regression Work Flow

It is a statistical system used for double bracket problems. It predicts the probability that a given input belongs to a particular class by using a logistic function. Although named retrogression, it's used for bracket tasks and can be extended to multiclass problems. Logistic retrogression is computationally effective but may underperform if the relationship between features is non-linear.

0	60	1	20	1	0	9	0
1	1	45	16	22	0	4	14
2	8	19	53	2	0	6	0
3	3	16	10	72	4	12	0
4	1	1	0	0	0	0	1
5	7	5	6	7	1	37	0
6	0	2	0	0	0	0	118
	0	1	2	3	4	5	6

Figure 3.1.2.10: Logistic Regression Confusion Matrix

	precision	recall	f1-score	support
0	0.75	0.66	0.70	91
1	0.51	0.44	0.47	102
2	0.50	0.60	0.55	88
3	0.69	0.62	0.65	117
4	0.00	0.00	0.00	3
5	0.54	0.59	0.56	63
6	0.89	0.98	0.93	120
accuracy			0.66	584
macro avg	0.55	0.56	0.55	584
weighted avg	0.66	0.66	0.66	584

Figure 3.1.2.11: Logistic Regression Values

3.1.3 Functional and Nonfunctional Requirements

The system must meet both functional and nonfunctional requirements to be effective and user-friendly.

Functional Requirements:

The functional requirements of the system include the ability to upload speech samples in various audio formats (e.g., WAV, MP3), followed by preprocessing steps such as noise removal, volume normalization, and speech segmentation. It must then

extract relevant acoustic features, such as Mel Frequency Cepstral Coefficients (MFCCs), pitch, tone, and rhythm, to distinguish between different regional accents. The system should support the training of machine learning models, including Naive Bayes, KNN, Random Forest, and Decision Tree, for accent classification. Upon receiving new speech samples, the system must classify them into predefined regional accents using the trained models. Additionally, it should provide performance evaluation metrics, including accuracy, precision, recall, and F1-score, to assess the model's effectiveness.

The system should classify new speech samples into the correct regional accent category, with the ability to output classification results to the user.

The system must include evaluation tools to assess the performance of the trained models, such as accuracy, precision, recall, and F1-score.

Nonfunctional Requirements:

Performance: The system should be optimized for fast data processing and classification. The model training and prediction times should be minimized through the use of efficient algorithms and hardware acceleration (e.g., GPU).

Scalability: The system must be scalable, allowing for the inclusion of additional regions or larger datasets in the future without significant modifications.

Reliability: The system should consistently provide accurate results and operate without crashes or errors, even with large datasets.

Usability: The system should have an intuitive user interface that is easy to navigate. Users should be able to upload speech samples, select models, and view results with minimal effort.

3.1.4 Context Diagram

A context diagram is a fundamental tool used to visually represent the boundaries and interactions of a system within its external environment. This diagram focuses on how the system communicates with external entities and how data is exchanged between them. In the case of the Speech-Based Classification of Bengali Regional Accents, the primary external entities interacting with the system are the **User** and the Speech Data. The User is responsible for uploading audio files to the system, selecting the classification algorithm, and reviewing the results provided by the system. These interactions are key to ensuring that the user is able to successfully classify the regional accents in the Bengali speech samples.

Once the User uploads speech samples, the system processes the audio by extracting relevant features, such as pitch, rhythm, and tone, which are distinctive to various Bengali regional accents. The machine learning model is then used to analyze these features and predict the regional accent of the speaker. After the classification process is complete, the system outputs the predicted accent, accompanied by performance metrics that can be reviewed by the user.

The Speech Data serves as the core input to the system, which is processed through several stages including preprocessing, feature extraction, model training, and

classification. This interaction between the system and the external entities is critical to ensure the successful operation of the accent classification system, and the context diagram visually maps these interactions for a better understanding of the overall system functionality. It highlights the data flow and the various stages involved, ensuring that each step is understood and executed correctly.

3.1.5 Data Flow Diagram Level 1

The DFD expands on the context diagram by offering a more detailed breakdown of how data moves through the system, from its initial input to its final output. The DFD starts with the raw speech data uploaded by the User. Once the data is received, it goes through a preprocessing stage, where any unwanted noise is filtered out, and the speech is normalized for consistent processing. Preprocessing also involves segmenting the speech into smaller, manageable parts that can be efficiently analyzed.

After preprocessing, the data is passed to the feature extraction module, where critical features such as MFCCs, pitch, and rhythm are extracted. These features serve as the basis for classification, as they contain the essential phonetic and prosodic information needed to distinguish between different Bengali regional accents. The system then moves on to the machine learning module, where various models, such as Random Forest, Decision Trees, and KNN, are used to classify the speech based on the extracted features. Each model is trained on a labeled dataset and fine-tuned to ensure it can accurately predict regional accents.

Once the models have classified the speech data, the output module presents the results to the user. These results include the predicted regional accent, as well as performance metrics such as accuracy, precision, recall, and F1 score. This detailed flow ensures that the system performs efficiently and accurately, providing clear insights into the regional accents present in the speech samples. The DFD visually represents how data is processed and helps identify potential areas for improvement in the workflow.

3.1.6 UI Design

The **User Interface (UI)** design is an integral aspect of the system's success. A well-designed interface ensures that the system is accessible, user-friendly, and efficient, enabling users to interact with the system seamlessly. The UI allows users to easily upload speech data for classification, select the desired machine learning model, and view the classification results and performance metrics. The design focuses on simplicity and clarity, making the user experience as intuitive as possible.

Key features of the UI include a centralized upload button, which allows users to quickly submit speech samples for processing. Once the audio files are uploaded, a progress bar is displayed to indicate the status of the data processing. The UI also provides a clear area for displaying results, including the predicted regional accent and associated metrics. Visual representations of the model's performance, such as

accuracy charts or confusion matrices, allow users to evaluate how well the system performed and to identify any areas where the model could be improved.

Furthermore, the UI is designed with user guidance in mind. Clear instructions and tooltips are available to assist users in navigating the system and understanding each stage of the classification process. To enhance usability, the interface is optimized for various devices, ensuring that it is responsive on both desktop and mobile platforms. The design aims to provide an engaging, efficient, and informative experience for the users, contributing to the overall success of the system.

3.2 Detailed Methodology and Design

The detailed methodology outlines the entire process of developing the speech-based accent classification system, from initial data collection to the final evaluation of the model's performance. It provides a comprehensive roadmap that guides each phase of the project to ensure that the system is both functional and efficient. The methodology begins with data collection, which involves gathering a diverse set of Bengali speech samples from various regional accents. This dataset is crucial for training the system, as it ensures that the system is exposed to a wide range of accents and phonetic variations.

The preprocessing phase follows data collection, where speech samples are cleaned and normalized to remove background noise and standardize the format of the data. This ensures that the features extracted from the speech data are consistent and reliable. The feature extraction step is performed using techniques like MFCCs, which are essential for capturing the acoustic features that distinguish different accents. These features are then used as inputs for training machine learning models, such as Random Forest, Decision Trees, and K-Nearest Neighbors (KNN).

After training, the system is tested and evaluated using a validation dataset to ensure that the model can accurately classify the regional accents. The evaluation process includes the use of performance metrics such as accuracy, precision, recall, and F1 score, which provide a comprehensive understanding of the model's ability to correctly classify the speech data. The methodology also includes system optimization, where the model's parameters are fine-tuned to improve its performance, and scalability considerations are made to ensure that the system can handle larger datasets in the future.

3.3 Project Plan

The project plan provides a structured timeline of activities and milestones necessary to complete the research. The plan breaks the project down into several key stages:

Data collection: Gathering a diverse set of Bengali speech samples.

System design: Planning and designing the system architecture and data flow.

Model training: Training machine learning models on the preprocessed dataset.

Evaluation: Testing the system's performance and refining it based on evaluation metrics.

Final report writing: Documenting the results, analysis, and conclusions of the study.

Each stage is accompanied by specific deliverables, and the project plan ensures the timely completion of all phases.

3.4 Task Allocation

Effective task allocation is essential for organizing the workflow and ensuring that each aspect of the project is addressed efficiently. In this project, tasks are divided into key phases such as data preprocessing, feature extraction, model training, evaluation, and user interface design. Depending on the team size, each team member is assigned specific responsibilities based on their expertise and experience. For example, one team member might focus on data preprocessing and feature extraction, while another might handle the model training and evaluation.

This clear division of tasks helps streamline the development process, ensuring that no part of the project is overlooked and that all tasks are completed on time. Regular progress meetings and reviews help ensure that the team is on track and that any issues or challenges are addressed promptly.

3.5 Summary

The research methodology serves as a comprehensive and structured blueprint for the design and development of the speech-based classification system for Bengali regional accents. This methodology meticulously outlines each phase of the project, ensuring a clear path from inception to completion. The process begins with data collection, where diverse speech samples from various Bengali regional accents are gathered. This diversity of data is critical for ensuring that the system is capable of recognizing and classifying a wide range of regional dialects accurately. The next phase is preprocessing, where the raw speech data is cleaned and normalized to remove noise and irrelevant information, making the data more suitable for analysis. Following preprocessing, **feature extraction** takes place, which focuses on identifying key features of the speech, such as pitch, rhythm, and Mel-Frequency Cepstral Coefficients (MFCCs), which are crucial for differentiating between accents.

Once the features are extracted, the system moves into the model training phase. Here, various machine learning algorithms, such as Random Forest, Decision Trees, and K-Nearest Neighbors (KNN), are applied to the features to build a model capable of classifying the accents. The model is then evaluated using a set of performance metrics, such as accuracy, precision, recall, and F1 score, to assess how well it can classify unseen data and generalize to different speech samples. The methodology also prioritizes user experience by designing a user-friendly interface that allows users to easily upload speech samples, view results, and access performance metrics.

This ensures that the system is not only technically sound but also accessible to users with varying levels of technical expertise.

Additionally, the project plan and task allocation strategies are designed to ensure that the research progresses efficiently and on time. Clear timelines and specific milestones ensure that each task is completed within the designated period. The evaluation metrics provide a solid framework for assessing the accuracy and effectiveness of the model, while regular performance reviews ensure that the system remains aligned with its goals. By adhering to this detailed methodology, the project ensures that the system will be capable of accurately classifying Bengali regional accents, offering both technical reliability and an optimized user experience.

In summary, this research methodology represents a well-organized, systematic approach to developing the speech-based classification system, ensuring that the project is carried out effectively and efficiently. Each phase of the methodology—from data collection to model evaluation and system design—works in tandem to create a robust system. The combination of thorough data analysis, advanced machine learning techniques, and a focus on user accessibility guarantees the creation of a high-performing system that can handle the complexities of Bengali regional accents. Moreover, the methodology ensures that the project remains on track through effective project management and task allocation, while providing the means to rigorously assess the system's performance and identify areas for improvement. This holistic approach will ultimately lead to the successful realization of a tool that enhances speech recognition for Bengali speakers and contributes to the advancement of machine learning applications in linguistic diversity.

Chapter 4

Implementation and Results

The implementation phase of the project focuses on developing the speech-based classification system for Bengali regional accents, integrating all the components defined in the design specification. This phase involved setting up the required software and hardware environment, developing the necessary modules, and conducting thorough testing to ensure the system's reliability.

4.1 Environment Setup

For the implementation of the speech-based classification system, we utilized a cloud-based environment to streamline data access, processing, and model training. This environment provided an easy-to-use interface that allowed for the seamless execution of the required machine learning code and the efficient management of large datasets. The specific tools used in this setup were Google Colab and Google Drive.

Google Colab was chosen as the primary platform for coding and training machine learning models due to its convenience and powerful features. It provides a fully integrated Jupyter notebook interface, where Python code can be executed in real-time, allowing for immediate feedback and adjustments. One of the major advantages of using Google Colab is its support for GPU and TPU acceleration, which significantly improved the training time for computationally expensive models, such as Random Forests and Decision Trees, by utilizing the hardware resources provided by Google.

Google Drive served as the storage solution for the large speech dataset used in this study. By storing the dataset in Google Drive, we were able to easily share and access the data between different instances of Colab, ensuring that the data was securely backed up and available throughout the entire implementation process. Google Drive's integration with Google Colab allowed for straightforward mounting of the drive within the notebook environment, facilitating direct reading and writing of data files.

In addition to these primary tools, various Python libraries were used to handle the data and implement the machine learning models. Libraries such as NumPy, Pandas, and Scikit-learn were essential for data manipulation, model building, and evaluation. The ease of setting up and running the environment on Google Colab,

combined with the file management provided by Google Drive, made this cloud-based setup an ideal choice for the project.

Overall, the choice of Google Colab and Google Drive enabled a flexible, efficient, and scalable environment for conducting machine learning experiments, which contributed to the success of the implementation phase of this research.

4.2 Comparative Analysis

A detailed comparative analysis was conducted to benchmark the performance of the proposed Random Forest algorithm, which achieved the highest accuracy in this study, against algorithms from previous research in speech-based classification systems. This analysis evaluates accuracy, dataset size, language focus, and methodology, providing a holistic view of the advancements made.

Table 4.2: Comparative Analysis

Study	Algorithm	Dataset	Accuracy (%)
Current Study	Random Forest	Bengali regional accents	98.12%
Accent Recognition Study	svm	Machine English accents	92.45%
Speech Recognition Study	Decision tree	Mixed regional languages	87.67%
Accent Detection Study	Naive Bayes	Multilingual speech samples	84.24%

4.3 Results and Discussion

The results of this study indicate that machine learning algorithms can indeed be effective in classifying Bengali regional accents based on speech data, with varying degrees of success depending on the chosen algorithm. The primary objective of the testing phase was to assess the classification accuracy of different machine learning models, and based on the results, we can make several observations and conclusions about the models' performance.

The Random Forest algorithm performed the best, achieving an impressive classification accuracy of 98.12%. This result suggests that the Random Forest model was able to effectively capture the complex, non-linear relationships present in the speech data. The Random Forest algorithm works by constructing multiple decision trees and combining their outputs, which helps to reduce the risk of overfitting and improves the overall generalizability of the model. This is particularly important in

tasks like regional accent classification, where speech patterns may vary widely across different accents.

On the other hand, the Decision Tree model, while still performing well, achieved a lower accuracy of 87.67%. The Decision Tree algorithm builds a single tree to classify the data, which makes it more prone to overfitting, especially when dealing with complex and noisy datasets like speech. Despite this, it still provided relatively good results, but improvements could be made by incorporating ensemble methods, such as Random Forest, which may overcome the limitations of a single tree.

The KNN algorithm also showed a decent performance with an accuracy of 75.17%. KNN is a simple yet effective algorithm, but it tends to struggle when the dataset is large or when the feature space is highly dimensional. In this case, the KNN model might not have been able to adequately distinguish between the subtle variations in the speech features across different Bengali regional accents. As KNN relies heavily on distance metrics, its performance can be sensitive to the feature scaling and the choice of distance metric used, which could explain the lower accuracy in this case.

Finally, the Logistic Regression model achieved the lowest accuracy of 65.92%. Logistic Regression is a linear classifier, meaning that it assumes a linear relationship between the input features and the target variable. However, speech data often exhibits complex and non-linear patterns, making it a challenging task for Logistic Regression to capture. While Logistic Regression is a fast and interpretable model, it may not be suitable for tasks where the underlying data is non-linear, as in the case of regional accent classification.

In terms of the comparative analysis, it is evident that ensemble methods, such as Random Forest, are highly effective for this task, given their ability to handle complex patterns and reduce overfitting. Additionally, models like Decision Tree and KNN could benefit from further tuning or hybrid approaches, while Logistic Regression may not be the most suitable for this type of classification task.

Moreover, the results also suggest that feature engineering and the inclusion of more advanced techniques, such as deep learning models, could potentially improve the performance further. Given that speech data is highly complex and rich in information, leveraging deep learning architectures like convolutional neural networks (CNNs) or recurrent neural networks (RNNs) could provide additional improvements in classification accuracy.

4.4 Summary

In conclusion, the implementation and evaluation of machine learning models for the classification of Bengali regional accents based on speech data demonstrated that machine learning techniques are highly effective for this task, with varying degrees of success depending on the model used. The Random Forest algorithm stood out as the most effective approach, achieving an accuracy of 98.12%, which underscores its strength in handling complex, non-linear relationships and its ability to generalize well to unseen data.

While the Decision Tree, K-Nearest Neighbors (KNN), and Logistic Regression models also showed reasonable performance, they fell short compared to Random Forest, particularly in handling the diversity and complexity inherent in the speech data. The results highlight the importance of model selection and tuning when dealing with challenging classification tasks such as regional accent detection.

This study has provided valuable insights into the feasibility of using machine learning for regional accent classification and has laid the groundwork for future improvements in this area. Further research could explore the potential for incorporating more advanced models, such as deep learning algorithms, as well as the role of feature extraction techniques in enhancing model performance.

Additionally, as the dataset used in this study was relatively small and specific to certain regions of Bangladesh, expanding the dataset to include more diverse samples from other regions or countries could provide even more robust results and generalizability.

Ultimately, the findings from this research offer a solid foundation for the development of speech-based systems capable of accurately classifying regional accents, with applications in areas such as speech recognition, automated transcription, and language processing.

Chapter 5

Engineering Standards and Design Challenges

In the development of the "Speech-Based Classification of Bengali Regional Accents using Machine Learning" system, adherence to relevant engineering standards is essential to ensure that the system is both reliable and effective. These standards guide the design and implementation process, ensuring that the system functions as intended and meets the expected quality criteria.

5.1 Compliance with the Standards

Compliance with engineering standards is essential for ensuring reliability, interoperability, and safety in the development of technological systems. Adherence to established standards facilitates compatibility across devices and systems while maintaining quality and performance. For speech-based classification of Bengali regional accents, the integration of software, hardware, and communication standards is critical to achieving robust and scalable solutions.

By following recognized industry standards, this project aims to enhance system reliability and user trust. Standards not only provide guidelines for implementation but also help mitigate risks related to technical failures and security vulnerabilities. This section explores the relevant standards in software, hardware, and communication, emphasizing their importance in the context of this research.

5.1.1 Software Standards

Software standards play a pivotal role in the development of reliable and maintainable systems. For this project, compliance with international standards like ISO/IEC 25010 for software quality is crucial. This standard ensures that the software meets requirements for functionality, reliability, usability, and efficiency. Additionally, adherence to IEEE 12207 for software lifecycle processes helps maintain consistency and quality across all development phases, from design to deployment.

In the context of machine learning models used for accent classification, software standards guide the implementation of algorithms and data processing frameworks. Using standardized programming languages such as Python and libraries like TensorFlow or PyTorch ensures compatibility and ease of integration. Moreover, standards for data exchange, such as JSON or XML, facilitate seamless communication between software components.

Another key consideration is data security. Adhering to standards like the General Data Protection Regulation (GDPR) ensures the ethical handling of user data. This is particularly important for datasets containing speech recordings, as they may involve sensitive personal information. Compliance with these standards not only protects user privacy but also enhances the credibility of the system.

5.1.2 Hardware Standards

Hardware standards ensure that physical components of a system are reliable, interoperable, and meet performance expectations. For the speech classification system, hardware components like microphones, servers, and data storage devices must comply with established standards to guarantee accurate data capture and processing.

Microphones and recording devices must adhere to standards such as IEC 61672 for sound level meters, ensuring accurate and consistent audio input quality. This is critical for capturing the subtle acoustic variations that distinguish Bengali accents. Additionally, compliance with standards like IEEE 802.3 for Ethernet ensures reliable and high-speed data transfer during training and testing phases.

Server infrastructure plays a crucial role in handling the computational demands of machine learning models. Standards like ISO/IEC 27001 for information security management systems provide guidelines for securing hardware against unauthorized access and cyber threats. Furthermore, compliance with energy efficiency standards such as ENERGY STAR helps minimize the environmental impact of the hardware used in the project.

Storage devices must also meet performance and reliability standards, such as those outlined in ISO/IEC 11801 for structured cabling systems. These standards ensure that the storage solutions can handle large datasets and facilitate smooth retrieval of data during model training and validation.

5.1.3 Communication Standards

Communication standards are integral to the seamless exchange of data between various components of the system. For a machine learning-based classification system, adherence to communication protocols and standards ensures efficient and error-free data transmission.

Standards like IEEE 802.11 for wireless communication provide the foundation for connecting devices in a networked environment. These protocols ensure that the

system can transmit large datasets, including speech recordings and model outputs, without interruptions or data loss. Similarly, the use of TCP/IP protocols ensures reliable and secure communication over the internet.

For interoperability between devices and systems, adherence to open communication standards such as RESTful APIs is essential. These standards enable different software components to communicate effectively, facilitating smooth integration with external systems like cloud platforms and mobile applications. Additionally, compliance with standards like HTTPS ensures secure data exchange, protecting sensitive information from unauthorized access.

In the context of speech processing, communication standards also include protocols for encoding and transmitting audio data. Standards such as WAV or FLAC for audio formats ensure that the recordings retain high fidelity during transmission and storage. These formats are particularly suited for analyzing acoustic features critical to accent classification.

By adhering to these communication standards, the project ensures that the system remains compatible with existing technologies and can be easily scaled or integrated into larger frameworks. This compliance not only enhances system performance but also ensures its long-term viability in diverse applications.

5.2 Impact on Society, Environment and Sustainability

The implementation of speech-based classification of Bengali regional accents holds transformative potential across numerous aspects of life, improving both the everyday experiences of individuals and the functioning of essential services. For individuals, the key benefit is its ability to address a significant gap in digital accessibility. Traditional speech recognition systems have primarily been optimized for standard accents, which creates barriers for individuals who speak with regional accents. This gap can lead to frustration, reduced usability, and a sense of exclusion when interacting with digital platforms, such as virtual assistants, voice-activated devices, and customer support systems. The development of a system that accommodates the diverse regional accents of Bengali ensures that users are no longer marginalized based on their accent, fostering inclusivity and enhancing their ability to engage with technology effortlessly.

5.2.1 Impact on Life

Beyond individual user experiences, this system has the potential to revolutionize the education sector. Language learning, particularly for non-native speakers, can benefit from personalized feedback on pronunciation and accent adaptation. For students learning Bengali as a second language or for those in multilingual regions trying to master specific regional dialects, the system can offer tailored pronunciation exercises, providing an interactive and immersive learning experience. In addition, this system can help students who want to refine their understanding of Bengali's regional dialects by accurately identifying pronunciation patterns and offering corrective feedback, promoting a deeper appreciation for the diversity of the language. By using technology to better understand regional

variations in accent, the system encourages cultural exploration and cross-cultural appreciation, which is invaluable in multicultural and multilingual societies.

The healthcare industry stands to gain immensely from this innovation, as clear communication is crucial for delivering effective care. Voice-activated systems designed to accommodate regional accents can significantly enhance telemedicine services, improving communication between healthcare providers and patients. In medical scenarios, especially in emergencies where every second counts, accurate and reliable communication can be the difference between life and death. By ensuring that healthcare systems can understand and process the diverse accents of Bengali speakers, the technology will help reduce misunderstandings and facilitate timely, precise medical interventions. The ability to recognize and interpret regional accents will be particularly beneficial in rural or underserved areas, where the population may predominantly speak with a specific regional accent.

The project also serves a significant cultural purpose. By documenting and analyzing Bengali regional accents, this research will establish a linguistic archive that preserves the rich phonetic diversity of the Bengali language. As languages evolve and accents change over time, it becomes essential to capture and preserve the linguistic features that define them. This archive will not only serve as a valuable resource for linguistic studies, but it will also ensure that these accents are not lost as Bengali continues to evolve in the face of globalization. It will be an essential tool for future generations of linguists, educators, and cultural researchers who seek to understand the historical and sociolinguistic context of regional Bengali accents.

5.2.2 Impact on Society & Environment

The societal impact of this technology is profound. By enabling more accurate recognition of regional accents, the system promotes linguistic inclusivity, ensuring that people from different regions can interact with digital platforms without facing discrimination or frustration. The system reduces digital exclusion by accommodating underserved communities whose accents have historically been underrepresented in speech recognition technologies. It also champions linguistic diversity, fostering a sense of pride and respect for regional differences. By acknowledging and celebrating the diverse ways people speak, the project encourages mutual understanding, helps bridge cultural divides, and strengthens social cohesion. As people from different regions are able to communicate seamlessly through voice-activated systems, the technology fosters a sense of belonging, particularly for marginalized communities who may have felt alienated in mainstream technological environments.

On an environmental level, this project is designed with sustainability at its core. Machine learning models will be optimized to improve computational efficiency, which in turn minimizes the environmental impact of both the training process and the usage of the technology. The adoption of cloud-based platforms powered by renewable energy sources further strengthens the project's commitment to environmentally sustainable practices. Additionally, the ethical considerations embedded in the data collection and model development process help ensure that the

project minimizes unnecessary resource consumption, contributing to a more sustainable future in technology development.

The technology also has the potential to drive economic growth and development. By improving regional voice recognition, local businesses can enhance customer service experiences through more efficient, voice-enabled customer support systems. This would result in higher customer satisfaction and loyalty, benefiting both businesses and consumers, especially in regions that have traditionally been underserved by the major players in the tech industry. Moreover, in the educational sector, this system provides equal learning opportunities for students in remote areas, reducing educational disparities and promoting equal access to quality resources. In this way, the project not only benefits individuals but also plays a role in the economic development of local communities, helping them to become active participants in the broader digital economy.

5.2.3 Ethical Aspects

Ethics plays a central role in guiding the development and implementation of this project, ensuring that it adheres to the highest standards of fairness, privacy, and responsibility. Protecting user privacy is a core principle, with strict protocols in place to anonymize and securely store speech data. All speech recordings used in the model training process are anonymized to safeguard personal identities, and the project complies fully with data protection regulations such as the General Data Protection Regulation (GDPR). By adhering to these standards, the project fosters trust among users, ensuring that they feel confident about participating in the system without compromising their privacy.

Addressing issues related to representation bias is another key ethical priority. It is vital to ensure that the dataset used to train the model includes a broad and diverse range of regional accents. A balanced representation of accents ensures that the system's performance remains equitable across all groups, avoiding the risk of marginalizing certain communities. Additionally, transparency in the algorithmic decision-making process is essential. By offering users insight into how decisions are made and how their data is used, the project can build greater trust and accountability.

The potential for misuse of this technology is also a concern. As with any AI-driven system, it is crucial to establish clear policies and ethical guidelines to prevent the technology from being used unethically, such as in surveillance or unauthorized profiling. By addressing these risks upfront and ensuring that the system's applications remain aligned with societal values of inclusivity and respect for individual rights, the project safeguards its integrity. These ethical considerations are woven into every phase of the project, ensuring that it remains a positive force in society.

5.2.4 Sustainability Plan

The sustainability plan for this project is multi-faceted, focusing on environmental, economic, and social sustainability. From an environmental standpoint, the project

employs energy-efficient machine learning algorithms and hardware to minimize its power consumption during training and inference stages. Additionally, the project makes use of cloud-based infrastructure powered by renewable energy sources, which aligns with global sustainability initiatives and helps reduce the project's overall carbon footprint.

Economically, the project aims to be both scalable and cost-effective. By leveraging open-source tools and frameworks, the development costs are minimized, ensuring that the technology remains accessible to a wide array of users, from academic institutions to small businesses. Building partnerships with government bodies, academic institutions, and private-sector stakeholders ensures that there is ongoing support for the project, creating avenues for future growth and development. These collaborations also offer the potential for commercialization, which can drive economic growth and create new opportunities for innovation in the tech industry.

Social sustainability is a critical aspect of the project. To ensure that the technology is well-received and effectively addresses the needs of different communities, the project emphasizes community engagement. This includes organizing workshops, awareness campaigns, and collaborating with local linguists, educators, and cultural organizations to ensure that the technology aligns with the values and needs of the regions it serves. By involving the community at every step of the project, the system's adoption is enhanced, and it becomes a tool for both cultural preservation and empowerment.

In conclusion, the sustainability plan ensures that the project's impact will be long-lasting and that it will continue to provide value in the future. By balancing environmental, economic, and social considerations, the project is positioned to deliver lasting benefits while fostering a harmonious integration of technology with societal needs and environmental priorities.

5.3 Project Management and Financial Analysis

Effective project management and financial analysis are critical to the success of the speech-based classification of Bengali regional accents. A structured approach ensures the project adheres to timelines, stays within budget, and achieves its objectives efficiently. This section outlines the key components of project management and provides a financial overview.

Project Management

The project management strategy is based on the principles of Agile methodology, allowing flexibility and iterative progress. The development lifecycle is divided into several phases: planning, data collection, model development, testing, deployment, and maintenance. Each phase has clearly defined deliverables, timelines, and roles assigned to team members to ensure accountability and seamless workflow.

A cross-functional team comprising linguists, machine learning engineers, software developers, and domain experts collaborates to meet project goals. Regular meetings and progress reviews ensure transparency and foster a collaborative environment. Risk management is a key focus, with contingency plans in place to address challenges such as delays in data collection or model training inefficiencies.

The use of project management tools like Trello and Jira helps in tracking progress and maintaining documentation. These tools also enable the team to identify bottlenecks early and take corrective actions, thereby enhancing productivity and efficiency.

Financial Analysis

The financial analysis involves estimating costs associated with data collection, infrastructure, personnel, and software tools. A significant portion of the budget is allocated to data collection and preprocessing, as this forms the foundation for model development. Costs for acquiring high-quality recordings and labeling data are carefully planned to ensure accuracy and diversity in the dataset.

SN	Components	Cost (BDT)
01	Data Collection	6000
02	Transport & Others	10000
	Total Estimated Cost	16000

In summary, robust project management practices combined with a well-structured financial plan ensure the project's success, scalability, and sustainability. This holistic approach balances technical, operational, and economic considerations to deliver impactful and lasting outcomes.

5.4 Complex Engineering Problem

The speech-based classification of Bengali regional accents using machine learning (ML) presents a multifaceted and complex problem, requiring the integration of various technical, linguistic, and ethical considerations. The first challenge lies in the inherent diversity of Bengali accents. As a language spoken across diverse geographical regions, Bengali exhibits significant variation in phonetics, tone, and pronunciation, influenced by historical, cultural, and social factors. These variations must be accurately captured to ensure that a machine learning model can classify accents with high precision. Achieving this requires large, diverse datasets containing speech samples from a wide range of speakers, each with distinct regional accents, and ensuring that these datasets are annotated and balanced to avoid bias.

5.4.1 Complex Problem Solving

A second challenge is the extraction of relevant features from speech data. Accents are manifested through subtle variations in acoustic features such as pitch, rhythm, intonation, and pronunciation patterns, which can be difficult to identify and quantify. Feature extraction techniques such as Mel-frequency cepstral coefficients (MFCCs) and prosodic analysis are essential but need to be finely tuned to capture the unique aspects of each regional accent. The performance of ML models depends heavily on how well these features are represented and processed.

Model development presents another complexity. Choosing the right machine learning algorithms is critical, as different algorithms perform differently based on the nature of the data. Supervised learning methods, such as support vector

machines (SVMs) and deep learning architectures like neural networks, need to be trained on annotated data. The models must be optimized to ensure accuracy, minimizing overfitting while achieving generalizability across new, unseen speech data.

Ethical challenges, including data privacy, representation bias, and the risk of misuse, must also be addressed. Ensuring the fair and responsible use of data, particularly in preserving linguistic diversity and preventing discriminatory outcomes, adds another layer of complexity to the problem.

Table 5.4.1: Mapping with complex problem solving.

EP	Dept of Knowledge	Range of Conflicting Requirements	Depth of Analysis	Familiarity of Issues	Extent of Applicable Codes	Extent of Stakeholder Involvement
EP 1	Data Science & Machine Learning	High: Trade-off between accuracy and real-time processing	Extensive: Involves feature extraction, model training, and validation	Moderate: Accents and regional variations are known but complex	High: Need for specialized algorithms and features (MFCC, pitch)	High: Researchers, linguists, and machine learning experts
EP 2	Signal Processing	Moderate: Need to balance quality and computational cost	Moderate: Requires extraction of meaningful features from speech data	High: Accents are well-known but the classification model is new	High: Well-defined algorithms for feature extraction (MFCC, pitch)	Moderate: Involves collaboration between speech engineers and data scientists

- **Rationale for EP2:** EP2 emphasizes signal processing in the classification of Bengali regional accents. There are conflicting requirements between accuracy and computational cost, as efficient models are required without sacrificing performance. The depth of analysis here focuses on the feature extraction process from speech data. Stakeholder involvement includes linguists, speech engineers, and machine

learning experts, emphasizing the interdependence between these disciplines for creating an effective model.

Table 5.4.2: Mapping with knowledge Profile.

K3	Engineering Fundamentals	K4	Specialist Knowledge	K5	Engineering Design	K6
EP1	Fundamental understanding of machine learning and signal processing	Advanced knowledge of regional dialects, phonetics, and language-specific characteristics	Expertise in designing machine learning models for speech processing	Practical knowledge of implementing speech recognition systems and handling challenges in real-world applications	In-depth research on accent classification methods, evaluating algorithms, and optimizing models	Engineering practice applies theoretical knowledge to design, test, and optimize machine learning models for accent classification.

5.4.2 Engineering Activities

The project encompasses a wide range of engineering activities, each tailored to tackle specific aspects of the complex problem. The initial phase involves designing a robust data collection framework to gather diverse speech samples representing Bengali regional accents. This includes recruiting participants, setting up recording protocols, and ensuring ethical compliance in data handling.

Data preprocessing is a crucial engineering task, involving cleaning, segmentation, and normalization of speech data. Feature extraction techniques, such as Mel-frequency cepstral coefficients (MFCCs) and spectrogram analysis, are applied to transform raw audio into meaningful inputs for machine learning models. These activities require specialized tools and expertise to ensure accuracy and reliability.

The development phase focuses on creating and fine-tuning machine learning models. Engineers experiment with various architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to achieve the best results. Hyperparameter tuning, model validation, and iterative testing are integral to this process.

Deployment engineering activities involve integrating the trained models into user-friendly applications. This includes developing application programming interfaces (APIs) and graphical user interfaces (GUIs) that facilitate interaction with end-users. Engineers also address hardware and software compatibility to ensure seamless operation across devices.

Finally, continuous monitoring and maintenance activities are performed post-deployment. These include analyzing system performance, addressing user feedback, and updating models to incorporate new data. Such activities are vital to maintaining the system's effectiveness and relevance.

5.5 Summary

The speech-based classification of Bengali regional accents is a multifaceted engineering endeavor that addresses complex problems and involves diverse activities. The problem-solving process is driven by innovative approaches to tackle challenges such as linguistic diversity, computational efficiency, and real-world adaptability. By leveraging advanced machine learning techniques, the project achieves a balance between accuracy, robustness, and scalability.

Engineering activities span the entire lifecycle of the project, from data collection and preprocessing to model development and deployment. These activities are meticulously planned and executed, ensuring that each phase contributes effectively to the project's goals. Specialized tools and expertise play a pivotal role in overcoming technical challenges and delivering a high-performance system.

The project's outcomes extend beyond technical achievements, impacting society by promoting linguistic inclusivity and enhancing access to technology for diverse populations. Ethical considerations, such as data privacy and representation bias, are integral to the project, ensuring that its benefits are equitable and sustainable.

In conclusion, the project exemplifies the application of engineering principles to address a complex problem with significant societal and cultural implications. Through a combination of innovative solutions, rigorous engineering activities, and ethical practices, the project not only meets its immediate objectives but also paves the way for future advancements in speech technology and cultural preservation.

Chapter 6

Conclusion

In conclusion, this study successfully applied machine learning to address a critical challenge in speech recognition. While Random Forest emerged as the most effective model, the results from Decision Tree, KNN, and Logistic Regression also provide valuable insights into the capabilities and limitations of different algorithms. The findings not only advance the field of speech-based accent classification but also highlight the scope for improving regional language technologies. By tackling accent diversity, this research contributes to bridging gaps in natural language processing for Bengali speakers. This achievement underscores the potential for scaling such methods to other linguistically diverse regions, reinforcing the global relevance of accent recognition technology.

6.1 Summary

This study delved into the classification of Bengali regional accents using speech data and machine learning techniques. Speech data was collected from seven regions, encompassing 580 audio samples from Chandpur, 535 from General Bengali, 484 from Bogura, 456 from Chittagong, 420 from Sylhet, 413 from Barishal, and 28 from other areas. The dataset underwent preprocessing to extract meaningful speech features, which were then used as inputs for machine learning models. Four algorithms were applied: Random Forest, Decision Tree, K-Nearest Neighbors (KNN), and Logistic Regression. Among these, Random Forest achieved the highest accuracy of 98.12%, showcasing its effectiveness in managing complex and diverse datasets. Decision Tree followed with 87.67%, while KNN and Logistic Regression achieved 75.17% and 65.92%, respectively.

The study underscores the capability of ensemble-based models like Random Forest in processing intricate speech data. The high performance of this model highlights its suitability for tasks involving linguistic diversity, such as accent recognition. This research also addresses the challenges of accent classification, including variability in speech patterns and imbalances in the dataset, which can affect model accuracy and generalizability. For instance, certain regions like “Others” were

underrepresented, posing challenges for comprehensive classification. Nonetheless, the findings contribute significantly to computational linguistics and accent recognition, offering a framework for leveraging machine learning to classify regional accents with high accuracy.

The study also emphasizes the importance of tailored approaches for accent classification. By focusing on speech patterns unique to Bengali accents, the research demonstrates how machine learning can cater to the specific linguistic needs of a region. Furthermore, the inclusion of multiple accents highlights the potential for creating more inclusive speech recognition systems. These systems can support applications such as automated transcription, accent-sensitive recommendations, and personalized language learning tools. The use of conventional machine learning methods provides a baseline for future work, paving the way for the incorporation of more advanced techniques.

6.2 Limitation

Despite its promising results, this study has several limitations. A significant issue is the imbalance in the dataset, with some regions like "Others" being severely underrepresented compared to others such as Chandpur and General Bengali. This imbalance may hinder the model's ability to generalize effectively across all regional accents. Additionally, while machine learning algorithms like Random Forest delivered high accuracy, they may not fully capture the complex temporal and contextual patterns in speech data. Another limitation is the study's focus on a limited number of regions, excluding accents from other Bengali-speaking areas, which restricts the generalizability of the findings. Finally, the use of traditional machine learning models, while effective, may not leverage the full potential of modern deep learning techniques that excel in speech-related tasks.

Another notable limitation is the reliance on manually annotated data, which introduces potential biases and inconsistencies in labeling. Furthermore, the study's evaluation metrics focused primarily on accuracy, without delving deeply into precision, recall, or F1-scores, which could provide a more nuanced understanding of model performance. The exclusion of environmental and contextual factors, such as variations in recording conditions or speaker emotions, may also impact the robustness of the models. Lastly, the computational efficiency of the employed algorithms was not analyzed, leaving room for uncertainty about their feasibility in real-time applications.

6.3 Future Work

Future research should address the limitations of this study by expanding the dataset to include more balanced and comprehensive samples from additional Bengali-speaking regions. This would ensure a more representative dataset and improve model performance across diverse accents. Exploring deep learning

methods, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), could enable more effective learning of the temporal and contextual patterns inherent in speech data. Additionally, integrating data augmentation techniques could help mitigate the issue of data imbalance and improve generalizability. Practical applications such as real-time accent recognition systems or mobile applications for regional language learning could also be developed to validate the model's performance in real-world settings. These advancements would further enhance the applicability and impact of speech-based accent classification systems.

Moreover, future work can explore the use of transfer learning by leveraging pre-trained models to accelerate training and improve performance. Multimodal approaches, combining speech with visual or textual data, could provide a more holistic understanding of regional accents. Incorporating socio-linguistic factors, such as age, gender, and social background, may also add depth to the analysis. Finally, collaboration with linguists and local communities could help create more context-aware and culturally sensitive systems, bridging the gap between technology and real-world linguistic diversity.

References

- [1] Ibrahim, Md Abu, Md Nawaz-S. Nayeem, and Sadaf Al Arabi. "Predicting regional accents of Bengali language using deep learning." PhD diss., Brac University, 2021.
- [2] Hassan, Muhammad Ahmed, Asim Rehmat, Muhammad Usman Ghani Khan, and Muhammad Haroon Yousaf. "Improvement in automatic speech recognition of south asian accent using transfer learning of deepspeech2." *Mathematical Problems in Engineering* 2022, no. 1 (2022): 6825555.
- [3] Rani, Sushmita, and Afroza Akhter Tina. "The impact of Bangla regional dialect on the pronunciation of English at tertiary level." *Humanities & Social Sciences Reviews*, eISSN (2020): 2395-6518.
- [4] Honnavalli, Dweepa, and S. S. Shylaja. "Supervised machine learning model for accent recognition in English speech using sequential MFCC features." In *International Conference on Artificial Intelligence and Data Engineering*, pp. 55-66. Singapore: Springer Nature Singapore, 2019.
- [5] Guntur, Radha Krishna, Krishnan Ramakrishnan, and Mittal Vinay Kumar. "An automated classification system based on regional accent." *Circuits, Systems, and Signal Processing* 41, no. 6 (2022): 3487-3507.
- [6] Guntur, Radha Krishna, Krishnan Ramakrishnan, and Vinay Kumar Mittal. "Automatic Classification of Foreign Language Accent." In *2021 2nd Global Conference for Advancement in Technology (GCAT)*, pp. 1-5. IEEE, 2021.
- [7] Sharma, Ch Rahul AN, Harsh Kumar Singh, H. Suhas Prabhu, Aniketh V. Jambha, C. Jyotsna, and Peeta Basa Pati. "Accent Detection in Indian Languages through Convolutional Neural Network based Spectrogram Analysis." In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1-5. IEEE, 2024.
- [8] Krishna, G. Radha, R. Krishnan, and V. K. Mittal. "An automated system for regional nativity identification of indian speakers from english speech." In *2019 IEEE 16th India Council International Conference (INDICON)*, pp. 1-4. IEEE, 2019.
- [9] Parikh, Pratik, Ketaki Velhal, Sanika Potdar, Aayushi Sikligar, and Ruhina Karani. "English language accent classification and conversion using machine learning." In *Proceedings of the International Conference on Innovative Computing & Communications (ICICC)*. 2020.
- [10] Shigli, Ashok, Ibrahim Patel, and K. Srinivasa Rao. "Automatic dialect and

- accent speech recognition of South Indian English." *Int. J. Latest Trends Eng. Technol* (2018): 103-111.
- [11] Bachate, Ravindra Parshuram, and Ashok Sharma. "Automatic speech recognition systems for regional languages in India." *International Journal of Recent Technology and Engineering* 8, no. 2 (2019): 585-592.
- [12] Hansen, John HL, and Gang Liu. "Unsupervised accent classification for deep data fusion of accent and language information." *Speech Communication* 78 (2016): 19-33.
- [13] Paul, Bachchu, Somnath Bera, Rakesh Paul, and Santanu Phadikar. "Bengali spoken numerals recognition by MFCC and GMM technique." In *Advances in Electronics, Communication and Computing: Select Proceedings of ETAEERE 2020*, pp. 85-96. Springer Singapore, 2021.
- [14] Islam, Md Atiqul, and An-Nazmus Sakib. "Bangla dataset and MMFCC in text-dependent speaker identification." *Engineering & Applied Science Research* 46, no. 1 (2019).
- [15] Paul, Bachchu, Himadri Mukherjee, Santanu Phadikar, and Kaushik Roy. "MFCC-based Bangla vowel phoneme recognition from Micro clips." In *Intelligent Computing and Communication: Proceedings of 3rd ICICC 2019, Bangalore 3*, pp. 511-519. Springer Singapore, 2020.
- [16] Biswas, Mainak, Saif Rahaman, Ali Ahmadian, Kamalularifin Subari, and Pawan Kumar Singh. "Automatic spoken language identification using MFCC based time series features." *Multimedia Tools and Applications* 82, no. 7 (2023): 9565-9595.
- [17] Behrman, Elizabeth, Arijit Santra, Siladitya Sarkar, Prantik Roy, Ritika Yadav, Soumi Dutta, and Arijit Ghosal. "Dialect identification of the bengali." *Data Science and Data Analytics: Opportunities and Challenges* (2021): 357.
- [18] Hassan, Foyzul, Mohammed Rokibul Alam Kotwal, Md Mostafizur Rahman, Mohammad Nasiruddin, Md Abdul Latif, and Mohammad Nurul Huda. "Local feature or mel frequency cepstral coefficients-which one is better for mln-based bangla speech recognition?." In *Advances in Computing and Communications: First International Conference, ACC 2011, Kochi, India, July 22-24, 2011. Proceedings, Part II 1*, pp. 154-161. Springer Berlin Heidelberg, 2011.
- [19] Bhattacharjee, Utpal, and Kshirod Sarmah. "Language identification system using MFCC and prosodic features." In *2013 International Conference on Intelligent Systems and Signal Processing (ISSP)*, pp. 194-197. IEEE, 2013.

Turnitin Originality Report

Processed on: 11-Jan-2025 23:08 +06
ID: 2562537883
Word Count: 16251
Submitted: 1

salman By Most. Hasna Hena

Similarity by Source	
Similarity Index	20%
Internet Sources:	15%
Publications:	12%
Student Papers:	10%

3% match (student papers from 28-Dec-2024)
[Submitted to Daffodil International University on 2024-12-28](#)

1% match (student papers from 31-Mar-2018)
[Submitted to Daffodil International University on 2018-03-31](#)

1% match (publications)
[H.L. Gururaj, Francesco Flammini, S. Srividhya, M.L. Chavadevi, Sheba Selvam, "Computer Science Engineering", CRC Press, 2024](#)

1% match (publications)
[Dinesh Goyal, Bhanu Pratap, Sandeep Gupta, Saurabh Raj, Rekha Rani Agrawal, Indra Kishor, "Recent Advances in Sciences, Engineering, Information Technology & Management - Proceedings of the 6th International Conference "Convergence2024" Recent Advances in Sciences, Engineering, Information Technology & Management, April 24-25, 2024, Jaipur, India", CRC Press, 2025](#)

1% match (Internet from 24-Dec-2024)
https://eitca.org/wp-content/uploads/materials/eitca_exam_mat_eitca-ai-dlitf.pdf

< 1% match (student papers from 13-Dec-2024)
[Submitted to Daffodil International University on 2024-12-13](#)

< 1% match (student papers from 16-Jul-2024)
[Submitted to Daffodil International University on 2024-07-16](#)

< 1% match (student papers from 12-Jul-2024)
[Submitted to Daffodil International University on 2024-07-12](#)

< 1% match (Internet from 05-Aug-2023)
<http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/10430/23111.pdf?isAllowed=y&sequence=1>

< 1% match (Internet from 03-Jun-2024)
<http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/12479/26705.pdf?isAllowed=y&sequence=1>

< 1% match (Internet from 29-Jun-2023)
<http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/9258/21285.pdf?isAllowed=y&sequence=1>

< 1% match (Internet from 06-Aug-2022)
<http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/8291/191-15-12253.pdf?isAllowed=y&sequence=1>

< 1% match (Internet from 29-Jun-2023)
<http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/8867/181-15-10699.pdf?isAllowed=y&sequence=1>

< 1% match (Internet from 21-Jul-2023)
<http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/10180/22812.pdf?isAllowed=y&sequence=1>