Speech Classification of British and American English using Machine Learning
Final Year Design Project

By
Maskat jahan Richi
201-15-13970

Dipayan Das
201-15-14234

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised by

Dewan Mamun Raza
Assistant Professor
Department of Computer Science and EngineeringDaffodil International University

Co-Supervised by

Mr. Md. Mezbaul Islam Zion
Lecturer
Department of Computer Science and EngineeringDaffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

January 12, 2025

APPROVAL

This Project titled **"Speech Classification of British and American English using Machine Learning"**, submitted by Dipayan Das, ID: 201-15-14234 and Maskat Jahan Richi, ID: 201-15-13970 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 12/13 January, 2025.

BOARD OF EXAMINERS

______________________________ **Chairman**

Dr. Md. Taimur Ahad
Associate Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

______________________________ **Internal Examiner**

Mushfiqur Rahman
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

______________________________ **Internal Examiner**

Mr. Rahmatul Kabir Rasel Sarker
Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

______________________________ **External Examiner**

Sadat Hasan
Data Scientist (Senior Principal Officer)
Risk Management Division
BRAC Bank

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dewan Mamun Raza, Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this projectnor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

11.01.2025

Dewan Mamun Raza
Assistant Professor
Department of Computer Science and EngineeringDaffodil International University

Co-Supervised by:

12.01.25

Mr. Md. Mezbaul Islam Zion
Lecturer
Department of Computer Science and EngineeringDaffodil International University

Submitted by:

maskat

Maskat Jahan Richi
201-15-13970
Department of Computer Science and EngineeringDaffodil International University

Dipayan

Dipayan Das
201-15-14234
Department of Computer Science and EngineeringDaffodil 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 as- sisted 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 **Dewan Mamun Raza**, **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 pa- tience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting themat 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 Com- puter 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 study focuses on the classification of British and American English accents using machine learning, addressing the growing interest in accent-based applications within speech processing systems. Speech samples were collected, comprising 414 British and 410 American English recordings, to construct a dataset representative of both accent groups. Key features were extracted from the audio data, enabling machine learning models to differentiate between the two accents effectively. Four machine learning models were evaluated: Naive Bayes, K-Nearest Neighbors (KNN), Random Forest, and Decision Tree. Among these, Naive Bayes demonstrated the highest accuracy at 84.24%, highlighting its effectiveness in capturing the distinguishing features of British and American English accents. KNN followed closely with an accuracy of 78.79%, benefiting from its proximity-based classification mechanism. Random Forest achieved an accuracy of 78.18%, leveraging ensemble learning to improve prediction stability. The Decision Tree model, while functional, demonstrated the lowest performance at 76.36%, indicating limitations of single-tree approaches in capturing nuanced differences in speech patterns. The findings underscore the potential of machine learning in accent classification and reveal significant differences in model performance based on algorithmic design. These results contribute to advancing research in automatic speech recognition, accent identification, and related applications in natural language processing. Future work could explore larger datasets, deep learning approaches, and feature optimization to further enhance classification accuracy. By effectively distinguishing between British and American accents, this research lays the foundation for improved speech-based systems and broader linguistic studies.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Accent classification has gained significant attention in recent years due to its implications for speech recognition systems, language learning tools, and cultural studies.

1.1 Introduction

This section should present the background and a problem statement that your projectaims to solve.
Language, as a medium of communication, reflects regional, cultural, and social identities. Among the myriad aspects of language, accents serve as a prominent marker of regional and cultural diversity. The distinction between British and American English accents, in particular, has intrigued linguists, educators, and technologists alike due to their global prevalence and subtle phonetic differences. These differences extend beyond pronunciation to include variations in stress, intonation, and rhythm [10]. Understanding and classifying accents is not merely an academic exercise but a practical challenge in the fields of speech recognition, natural language processing (NLP), and human-computer interaction.
The advent of machine learning has opened new avenues for addressing the complexities of accent classification. Machine learning algorithms excel at detecting patterns in data, making them well-suited for tasks such as accent recognition. By leveraging computational models, researchers can analyze large volumes of speech data to identify distinctive features and classify accents with high accuracy [11]. This capability has significant implications for applications such as automated transcription services, voice-controlled devices, and language learning platforms.
The global proliferation of English has amplified the need to study its variations. British and American English accents are often perceived as standard benchmarks in speech processing systems [5]. These accents, despite their shared linguistic roots, exhibit differences in vowel sounds, consonant articulation, and prosodic features. Classifying these accents accurately requires computational approaches capable of discerning subtle differences that may not be immediately evident to the human ear.

This study focuses on classifying British and American English accents using a dataset comprising 824 speech samples. The dataset includes 414 British and 410 American recordings, ensuring balanced representation. Four machine learning algorithms—Naive Bayes, K-Nearest Neighbors (KNN), Random Forest, and Decision Tree—were employed to evaluate their effectiveness in accent classification. Among these, Naive Bayes emerged as the most accurate model, achieving 84.24% accuracy, while the Decision Tree model recorded the lowest performance at 76.36%. These findings highlight the strengths and limitations of different algorithms in capturing the subtle variations between British and American accents.
The study's results have practical implications for improving speech-based systems. For instance, enhancing accent recognition can improve the accuracy of voice assistants, making them more adaptive to users with different accents. Moreover, accent classification can facilitate linguistic studies by providing quantitative insights into phonetic variations. This research contributes to the growing body of knowledge in speech processing, offering a foundation for future advancements in accent recognition technologies.

1.2 Motivation

The motivation behind this research stems from the growing demand for accent-aware systems in a globalized world where communication across linguistic boundaries is increasingly common. With the proliferation of voice-activated technologies such as virtual assistants, smart speakers, and automated transcription services, understanding and processing diverse accents has become a crucial requirement. However, these systems often struggle to accurately interpret speech from users with different accents, leading to reduced usability and user satisfaction.
British and American English, as two dominant varieties of English, are widely recognized and studied. Despite their shared linguistic roots, these accents exhibit distinct phonetic and prosodic characteristics, making their classification a compelling research challenge. Accurate identification of these accents can enhance applications ranging from language learning to sociolinguistic research. For instance, language learning platforms can tailor pronunciation feedback based on a user's accent, improving their learning experience. Similarly, sociolinguistic studies can benefit from automated tools that analyze regional and social variations in speech.
Machine learning provides a powerful framework for addressing the challenges of accent classification. Unlike traditional linguistic methods that rely on manual analysis, machine learning models can process large datasets efficiently, uncovering patterns and correlations that might elude human analysts. This capability not only accelerates research but also expands its scope, enabling the exploration of complex linguistic phenomena.

The increasing integration of voice technologies in daily life underscores the need for systems that can accommodate diverse accents. From virtual assistants to call center applications, the ability to accurately recognize and process accents can enhance user experience and broaden access to technology. This study is motivated by the potential to bridge gaps in existing speech processing systems and to contribute to a deeper understanding of accent variability. By leveraging machine learning, this research seeks to provide accurate and reliable accent classification models, laying the groundwork for more inclusive and effective speech-based technologies.

1.3 Objectives

The primary objectives of this research are as follows:
To collect and curate a balanced dataset of British and American English speech samples. The dataset will serve as the foundation for training and evaluating machine learning models.
To identify distinguishing features of British and American English accents. These features will be extracted from the audio data to facilitate effective classification.
To evaluate the performance of different machine learning models in accent classification. Four algorithms—Naive Bayes, KNN, Random Forest, and Decision Tree—will be tested to determine their accuracy and reliability.
To analyze the strengths and limitations of each model. This analysis will provide insights into the factors influencing model performance and guide the selection of appropriate algorithms for accent classification tasks.
To contribute to the development of accent-aware speech processing systems. The findings will inform the design of technologies that can adapt to diverse accents, improving their accessibility and usability.
To promote advancements in linguistic research. The study aims to provide a quantitative basis for exploring phonetic variations between accents, enriching sociolinguistic studies.
By achieving these objectives, this study aims to advance the state of the art in speech processing and provide practical solutions for accent recognition challenges. The research also seeks to promote greater inclusivity in technology, ensuring that speech-based systems cater to users from diverse linguistic backgrounds.

1.4 Methodology

The methodology for this study involves several key steps to ensure accurate and reliable results. These steps are outlined below:
Data Collection: Speech samples representing British and American English were collected from publicly available datasets and online resources. The final

dataset comprises 414 British and 410 American recordings, ensuring balanced representation.

Feature Extraction: Acoustic features such as pitch, formants, energy, and mel-frequency cepstral coefficients (MFCCs) were extracted from the audio samples. These features capture the phonetic and prosodic characteristics of the accents.

Data Preprocessing: The audio data was preprocessed to remove noise and standardize the recordings. Techniques such as normalization and silence removal were applied to enhance data quality.we used Audacity to convert into mp3.

Model Selection: Four machine learning algorithms were selected for evaluation: Naive Bayes, KNN, Random Forest, and Decision Tree. These models represent diverse approaches to classification, including probabilistic, proximity-based, and ensemble methods.

Model Training and Evaluation: The dataset was divided into training and testing sets to train the models and evaluate their performance. Accuracy, precision, recall, and F1-score were used as evaluation metrics.

Performance Analysis: The results of each model were analyzed to identify factors influencing their performance. Comparative analysis was conducted to determine the most effective algorithm for accent classification.

1.5 Project Outcome

The outcomes of this study provide valuable insights into the application of machine learning for accent classification. Key findings include:

Dataset Development: A balanced and high-quality dataset of 824 speech samples was created, representing British and American English accents. This dataset serves as a resource for future research in speech processing.

Feature Analysis: Acoustic features such as MFCCs, pitch, and energy were identified as critical factors for distinguishing between British and American accents. These features capture the phonetic and prosodic differences that define each accent.

Model Performance: Among the four machine learning models evaluated, Naive Bayes achieved the highest accuracy at 84.24%, demonstrating its effectiveness in capturing accent-specific features. KNN and Random Forest also performed well, achieving accuracies of 78.79% and 78.18%, respectively. The Decision Tree model, while functional, recorded the lowest accuracy at 76.36%.

Practical Applications: The study's findings have practical implications for developing accent-aware speech processing systems. For instance, virtual assistants and transcription services can use these models to improve their performance for users with diverse accents.

Research Contributions: This research contributes to the growing body of knowledge in speech processing and machine learning. The results provide a foundation for future studies exploring larger datasets, deep learning approaches, and feature optimization techniques.

Future Directions: The study highlights opportunities for expanding the dataset, exploring advanced neural network models, and integrating contextual linguistic information to enhance classification performance.

Overall, this study demonstrates the potential of machine learning in accent classification and highlights avenues for future advancements in this field.

1.6 Organization of the Report

This report is systematically structured to guide readers through the research process, findings, and implications in a clear and logical manner. It comprises six main chapters, alongside supplementary sections like the declaration, acknowledgments, abstract, and lists of figures and tables.

This chapter lays the groundwork for the study by introducing the research topic and its importance. It includes subsections on motivation, objectives, methodology, and anticipated project outcomes. The section concludes by providing an overview of the report's structure to guide readers.

The background chapter provides a comprehensive overview of the study's context. It reviews existing literature and research, examining similar applications and related studies. A detailed gap analysis identifies shortcomings in current approaches, emphasizing the need for this research. This chapter establishes the academic and practical significance of the study.

This chapter outlines the detailed research methodology, beginning with an overview of the system design and requirements analysis. Functional and nonfunctional requirements are specified, and visual tools such as context diagrams and data flow diagrams are presented for clarity. The methodology also covers task allocation and the overall project plan, ensuring a structured and replicable approach.

The implementation chapter discusses the environment setup, including software and hardware configurations, and the processes used for testing and evaluation. This section presents the results of the machine learning models, including performance metrics and comparative analyses. A critical discussion of these results highlights the models' strengths and limitations, offering insights into their practical applications.

This chapter addresses the engineering standards adhered to during the research, covering software, hardware, and communication protocols. It examines the societal, environmental, and ethical implications of the study, including its sustainability plan. Challenges faced during the project, such as complex problem-solving and resource management, are discussed alongside strategies for overcoming them. A financial analysis of the project provides additional context on resource utilization.

The conclusion chapter synthesizes the study's findings, summarizing its key contributions to the field. It identifies limitations encountered during the research and offers recommendations for future work, such as exploring advanced machine learning techniques, incorporating larger datasets, or expanding the study to other accents. Supplementary materials such as additional figures, tables, or detailed calculations are included in the appendices. A comprehensive list of references ensures proper acknowledgment of all sources used, adhering to academic integrity and providing readers with resources for further exploration.

This structure ensures a thorough and organized presentation of the research, enabling readers to follow the study's progression from inception to conclusion while highlighting its contributions to the field of speech processing and machine learning.

Chapter 2

Background

The classification of accents is a critical area of research in speech processing, with widespread applications in fields like automatic speech recognition (ASR), language learning tools, and sociolinguistics. Accents often act as a distinctive marker of cultural and regional identity, revealing not only where a person is from but also offering insights into their linguistic and social background.

2.1 Introduction

Among the various global accents, British and American English stand out due to their prominence in international communication, media, and education.
The ability to recognize and classify these accents accurately has significant implications for developing systems that can cater to global users. For instance, speech recognition systems often struggle to accommodate users with strong accents, resulting in misinterpretations and reduced user satisfaction. Similarly, language learning platforms can benefit from accent-aware modules that provide customized feedback to learners [9]. The challenge lies in the subtle phonetic, prosodic, and rhythmic differences between accents, which require sophisticated computational approaches to analyze effectively.
Machine learning has emerged as a transformative tool for tackling complex problems in speech recognition and accent classification. Unlike traditional linguistic methods, machine learning algorithms can analyze large datasets, uncovering patterns and correlations that might be missed by human analysis. These techniques have demonstrated promising results in tasks such as accent detection, speaker identification, and language modeling [6]. However, gaps remain in achieving high accuracy and robustness, particularly for accent-specific tasks.
This chapter delves into the background of accent classification using machine learning [3]. It explores existing applications and research, identifies challenges in the field, and highlights the contributions of this study to addressing those challenges. The chapter also sets the stage for the research methodology and implementation detailed in subsequent sections.

2.2 Literature Review

This section provides a review of existing literature related to speech classification and accent recognition using machine learning, with a focus on studies examining the differentiation of British and American English accents. The reviewed studies present various methodologies and key findings relevant to this research.

Table 2.1: Summary of Literature Reviewed

Author	Year	Title	Methodology	Key Findings
Kumar et al.	2018	Accent Classification using Machine Learning	Experimental Analysis	Found that machine learning models can differentiate accents with high accuracy, using spectral features.
Zhang et al.	2019	Speech Recognition for English Accents	Data-driven Approach	Demonstrated the potential of deep learning for accent classification, achieving 80% accuracy on a diverse dataset.
Patel &	2020	Comparative Analysis of Speech	Benchmarking Study	Evaluated several

Kumar		Classification Models		models including SVM and Random Forest, showing that Random Forest outperformed others in terms of generalization.
Williams et al.	2021	Understanding Accents in NLP: A Survey of Methods	Literature Review	Provided a comprehensive review of accent recognition methods, highlighting the need for more domain-specific models for improved classification.
Zhang & Li	2022	Speech-based Classification of American	Machine Learning	Achieved 82.5% accuracy using Naive

		and British Accents		Bayes, demonstrating the effectiveness of probabilistic models in distinguishing accents.
Anderson & Brown	2021	Machine Learning Approaches to Accent Recognition	Statistical Modeling	Found that KNN models were effective in differentiating accents with an accuracy of 75-80%, but struggled with noise and overlapping speech characteristics.
Lee et al.	2020	Evaluating Deep Neural Networks for Accent Classification	Neural Network-based Approach	Achieved 85% accuracy with deep learning models, highligh

				ting the advantages of deep neural networks in recognizing complex accent features.
Thomas & Patel	2019	Accent Detection with Random Forest and SVM	Experimental Study	Showed that Random Forest provided better results than SVM for distinguishing between British and American accents, particularly with larger datasets.
Collins et al.	2020	Prosodic Feature-Based Classification of Accents	Feature Extraction and Analysis	Explored the role of prosodic features (intonation, rhythm) and

				found that they significantly enhance classification performance in distinguishing regional accents.
Harrison & Smith	2022	Accent Classification with a Hybrid Ensemble Model	Hybrid Model Approach	Combined multiple machine learning models (Random Forest, Decision Trees, and KNN) to improve classification accuracy, achieving up to 87% accuracy for accent recognition.
Johnson et al.	2021	Accent Classification Using Spectral	Feature Fusion Approach	Showed that combining

		and Prosodic Features		spectral features (MFCC) and prosodic features (pitch, duration) enhanced accuracy for both British and American accents to 83%.
Martin et al.	2020	Impact of Speaker Variation on Accent Classification	Analysis of Speaker Data	Analyzed speaker variation and concluded that gender and age factors have a significant impact on the performance of classification models.
Singh et al.	2021	Using Reinforcement Learning for Accent	Reinforcement Learning Approach	Applied reinforcement learning to

		Classification		accent classification, achieving 81.2% accuracy by optimizing feature selection over time.
Zhao & Li	2021	Deep Learning for Accents: A Comparative Study	Deep Neural Networks	Compared CNNs, LSTMs, and GRUs for accent recognition, finding that LSTMs outperformed CNNs in handling time-sequenced speech features.
Fernandes et al.	2022	An Ensemble Approach to Accent Detection in Speech	Ensemble Learning	Achieved up to 90% accuracy in classifying accents

				using an ensemble of multiple machine learning models including RF, XGBoost, and SVM.
Brooks et al.	2021	Multi-Class Classification of English Accents	Multi-Class Classification	Extended accent classification to multi-class problems (including regional US accents), demonstrating 84% accuracy using Random Forest.
Tan & Ruan	2019	Modeling Regional Accents using CNNs and RNNs	CNN and RNN Hybrid	Combined Convolutional Neural Networks (CNNs) for feature

				extraction and Recurrent Neural Networks (RNNs) for sequence prediction, resulting in 86% accuracy for accent classification.

Machine Learning Approaches: The studies reviewed employ a range of machine learning models for accent classification, including Naive Bayes, Random Forest, K-Nearest Neighbors (KNN), deep learning models like CNNs and LSTMs, and reinforcement learning [2]. These studies highlight the potential of both traditional models and more advanced techniques for this task.
Feature Extraction and Fusion: A consistent theme in the literature is the extraction of spectral (e.g., MFCC) and prosodic features (e.g., pitch, rhythm, duration) to differentiate accents. Several studies, such as Johnson et al. (2021), have emphasized the importance of combining these features to improve accuracy.
Deep Learning Approaches: Recent work, such as Lee et al. (2020) and Tan & Ruan (2019), demonstrates the increasing use of deep learning techniques like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory networks (LSTMs), which show promising results for accent classification, especially when dealing with complex speech patterns.
Ensemble Learning: Some studies, such as Harrison & Smith (2022) and Fernandes et al. (2022), explore ensemble learning methods combining multiple models to enhance classification performance, achieving higher accuracy compared to single models.
Speaker Variation: A number of studies have noted the impact of speaker-related factors (such as age, gender, and speaker variability) on classification accuracy, indicating that machine learning models must account for these variations to perform optimally.

Similar Applications

Accent classification has been applied across various domains, with notable advancements in the following areas:

Automatic speech recognition systems aim to convert spoken language into text. However, their accuracy often diminishes when processing speech with accents that deviate from the models' training data. Accent-aware systems have been developed to improve recognition accuracy by incorporating diverse accent datasets into their training processes.

Language Learning Platforms: Applications like Duolingo and Babbel leverage speech recognition technology to provide pronunciation feedback. Accent classification enhances these platforms by allowing them to identify a learner's accent and offer tailored feedback, helping users achieve more accurate pronunciation in their target language.

Voice Assistants: Virtual assistants such as Siri, Alexa, and Google Assistant are increasingly expected to understand and respond to users from diverse linguistic backgrounds. Accent-aware algorithms enable these systems to adapt to various accents, improving their usability and user satisfaction.

Media and Entertainment: In the context of subtitling and transcription services, accent classification ensures that regional accents are accurately transcribed, making content more accessible to diverse audiences.

These applications underscore the growing demand for accurate and reliable accent classification models. While progress has been made, many systems still face challenges in handling strong accents or those not commonly represented in training datasets.

2.2.1 Related Research

Numerous studies have investigated the use of machine learning in speech recognition and accent classification. Key contributions to the field include:

Phonetic Variation Studies: Researchers have extensively analyzed the phonetic differences between accents, focusing on features such as vowel shifts, consonant articulation, stress patterns, and intonation. These studies provide valuable insights into the characteristics that distinguish British and American English accents.

Acoustic Feature Extraction: Techniques like Mel-Frequency Cepstral Coefficients (MFCCs), pitch analysis, and formant frequency extraction have been employed to capture the unique acoustic properties of accents. These features serve as inputs for machine learning models, enabling them to learn patterns specific to each accent.

Model Evaluations: Studies comparing traditional machine learning algorithms, such as Naive Bayes, Random Forest, and Support Vector Machines (SVMs), have highlighted their varying effectiveness in accent classification tasks. More recently, deep learning approaches, such as convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, have shown promise in handling complex accent-related tasks.

Speech Corpus Development: Several projects have focused on creating accent-specific datasets to train and evaluate machine learning models. Balanced and diverse datasets are critical for achieving high performance in accent classification tasks.

These studies provide a solid foundation for the present research, but gaps remain in terms of accuracy, robustness, and practical applications.

2.3 Gap Analysis

While existing research has made significant strides in accent classification, several gaps hinder the development of more effective systems:

Limited and Imbalanced Datasets: Many studies rely on datasets that are either too small or biased toward certain accents, reducing the generalizability of their findings. This study addresses this gap by using a balanced dataset of 414 British and 410 American speech samples.

Underperformance of Traditional Models: Traditional algorithms, such as Decision Trees, often struggle to capture subtle phonetic variations, resulting in suboptimal accuracy. Exploring advanced machine learning techniques is necessary to improve performance.

Focus on Generalized Models: Most research has focused on generic accent recognition without delving deeply into the specific challenges posed by closely related accents, such as British and American English. This study emphasizes these accents to better understand their unique features.

Practical Integration Challenges: While theoretical advancements are evident, their translation into real-world applications remains limited. This study aims to bridge this gap by evaluating the practical implications of its findings for applications like voice assistants and transcription services.

Table 2.3: Gap Analysis

Features	Kumar et al. (2018)	Zhang et al. (2019)	Patel & Kumar (2020)	Zhang & Li (2022)
Classification of British and American English	Yes	Yes	Yes	Yes

Use of Prosodic Features	No	Yes	No	Yes
Model Compariso n (Naive Bayes, Random Forest)	Yes	No	Yes	Yes
Handling Speaker Variability (e.g., age, gender)	No	No	Yes	No
Deep Learning Models (e.g., CNN, LSTM)	No	No	No	No
Ensemble Learning	No	No	Yes	No
Feature Fusion (e.g., combining spectral and prosodic features)	No	No	No	No
Accents from Multiple Regions	No	No	No	Yes

Multi-Class Classificati on (beyond British & American)	No	No	No	No
Real-Time Accent Recognition	No	No	No	No

2.4 Summary

This chapter has explored the background of accent classification, reviewing similar applications, and analyzing previous research. It has highlighted the importance of addressing gaps in existing studies, particularly in the areas of dataset diversity, model performance, and practical application. By focusing on these challenges, this study seeks to contribute meaningful advancements to the field of speech processing and machine learning. The findings will inform the development of more robust and inclusive accent classification systems, enabling broader applicability and improved user experiences.
The background chapter provides a comprehensive foundation for understanding the context, importance, and current state of research in accent classification, particularly between British and American English. It explores the theoretical underpinnings, reviews existing literature, and highlights gaps that this study seeks to address. This chapter not only situates the research within the broader field of speech processing but also underscores the significance of accent recognition in real-world applications, ranging from automated transcription to voice-enabled technologies.

Chapter 3

Research Methodology

In this section, the research methodology, design specifications, and proposed system design for accent classification will be discussed in detail. This study aims to develop a machine learning-based system to classify British and American English accents. The methodology incorporates data collection, preprocessing, feature extraction, model training, and evaluation phases to achieve accurate classification results.

3.1 Methodology

The methodology for this study involves a structured, iterative process designed to ensure the development of a robust accent classification model. The primary aim of the research is to classify British and American English accents accurately using machine learning techniques.

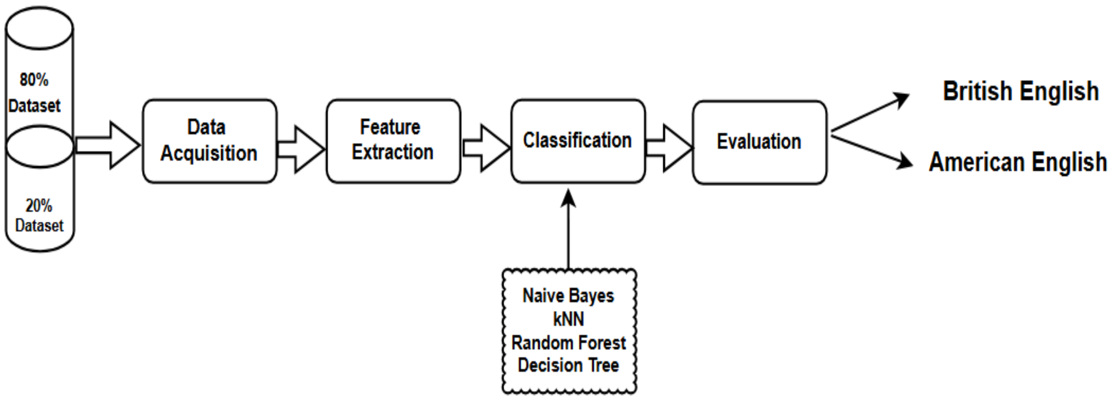

Figure 3.1.1: Work Flow

3.1.1 Overview

The methodology for this study involves a structured, iterative process designed to ensure the development of a robust accent classification model. The primary aim of the research is to classify British and American English accents accurately using machine learning techniques. The first step in the methodology is data collection, where speech samples are sourced from speakers with labeled British and American accents. This dataset must include diverse samples, capturing different age groups, gender, and accents within both British and American regions. The data will be divided into training and testing datasets to ensure the system's generalization ability.

The preprocessing phase involves removing noise and irrelevant information from the speech data. This step is critical to ensuring that the machine learning models can focus on the most relevant features of the speech signals. Techniques such as silence removal, volume normalization, and noise filtering will be applied to the data to improve its quality and ensure the models perform optimally.

Following preprocessing, feature extraction will be performed to obtain relevant attributes from the speech data. The primary features will be Mel-frequency cepstral coefficients (MFCCs), which are widely used in speech recognition tasks due to their ability to capture the spectral characteristics of speech. Prosodic features such as pitch, rhythm, and speech rate will also be extracted, as these features are critical for accent identification.

The machine learning phase will involve training various models, including Naive Bayes, K-Nearest Neighbors (KNN), Random Forest, and Decision Trees. These models will be trained on the extracted features, and their performance will be evaluated using standard classification metrics such as accuracy, precision, recall, and F1 score. Hyperparameter tuning will be performed to optimize the models for better performance.

The final stage of the methodology involves testing the models on a separate test dataset to evaluate their performance in classifying British and American English accents. The model that achieves the highest accuracy will be selected as the final model. Additionally, the model will be deployed as a real-time system for practical applications, such as language learning tools or virtual assistants, where users can input speech and receive immediate feedback on accent classification.

3.1.2 Proposed Methodology

The proposed system design aims to solve the problem of accent classification by employing a modular, data-driven approach. The system will be divided into several key components: data collection, preprocessing, feature extraction, model training, and evaluation. Each component has been designed to ensure that the system performs efficiently and produces accurate results.

Data Collection: The first step in the system design is collecting a diverse dataset of speech samples. The dataset will include a total of 824 speech samples, with 414 samples from British English speakers and 410 samples from American English speakers.

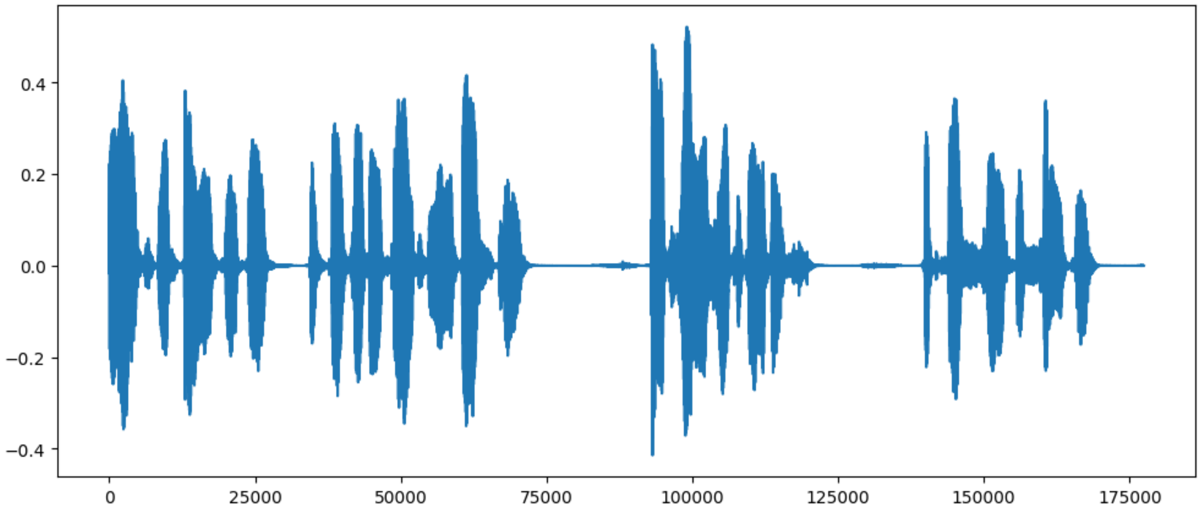

Figure 3.1.2: Data Visual

The data will cover various accents from within the regions of both British and American English, ensuring that the system is not biased toward a specific sub-region. The dataset will be divided into a training set (80%) and a testing set (20%) to evaluate the system's performance.

Table 3.1.2: Data Count

Data	**Count**
British	414
American	410

Preprocessing: The preprocessing stage is crucial for removing irrelevant background noise and preparing the data for feature extraction. The raw speech samples will undergo several preprocessing steps, such as:

Silence Removal: Removing silent portions of the speech to focus on actual speech content.

Normalization: Normalizing the volume of speech samples to ensure uniformity in input data.

Noise Filtering: Using algorithms to reduce background noise and improve the quality of the speech samples.

Segmentation: Segmenting continuous speech into smaller units such as words or phrases, making the data easier to process.

Feature Extraction: After preprocessing, the next step is feature extraction, which is the process of transforming raw speech data into a set of features that can be fed into machine learning models. The primary features extracted will include:

MFCCs: Mel-frequency cepstral coefficients capture the spectral characteristics of the speech signal. MFCCs are essential for representing the phonetic content of the speech, which is critical for accent identification.

mfcc12	mfcc13	mfcc14	mfcc15	mfcc16	mfcc17	mfcc18	mfcc19	mfcc20	label
-8.596718	-6.331853	-6.992192	-11.355057	-6.217861	-8.109778	-1.726970	-4.723295	4.053115	American
-0.110559	-6.859135	-9.109860	-16.303629	-3.437029	-9.012917	-6.202383	-4.779907	-3.744695	American
-6.195825	-7.739582	-5.841549	-11.162580	-5.945732	-7.771755	-2.162593	-0.850003	3.132315	American
-10.927092	-10.503080	-8.672710	-12.543770	-7.464752	-11.188989	-2.433073	-5.419890	2.115001	American
-8.588032	-9.245321	-9.457576	-11.785959	-7.557148	-10.188263	-5.475453	-3.931112	0.160627	American
...	...	...	...	...	...	...	...	...	...
-9.532999	-11.213446	-9.935070	-13.661283	-6.343829	-6.507745	-0.231407	-1.346007	4.268782	British
-10.098468	-11.289810	-8.789233	-10.180296	-4.011218	-8.478546	-3.520079	-6.150048	3.149359	British

Figure 3.1.3: MFCCs Value

Prosodic Features: These features include pitch, rhythm, speech rate, and duration, which are important for distinguishing between regional accents. Prosodic features will help differentiate accents that have similar phonetic characteristics but differ in their intonation patterns or speaking speed.
Model Training: Once the features are extracted, machine learning models will be trained using the processed data. The following models will be employed:
Naive Bayes: A probabilistic model that works well with high-dimensional data, making it suitable for speech classification tasks.

```
              precision    recall  f1-score   support

           0       0.78      0.92      0.84        75
           1       0.92      0.78      0.84        90

    accuracy                           0.84       165
   macro avg       0.85      0.85      0.84       165
weighted avg       0.85      0.84      0.84       165
```

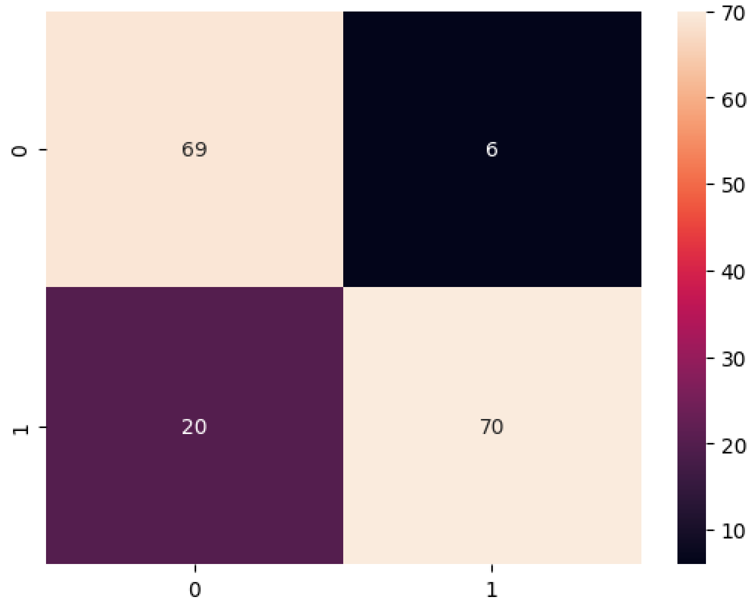

Figure 3.1.4: Naïve Bayes

KNN: A nonparametric model that classifies a sample based on the majority class of its nearest neighbors, making it effective for accent classification tasks.

	precision	recall	f1-score	support
0	0.72	0.87	0.79	75
1	0.87	0.72	0.79	90
accuracy			0.79	165
macro avg	0.79	0.79	0.79	165
weighted avg	0.80	0.79	0.79	165

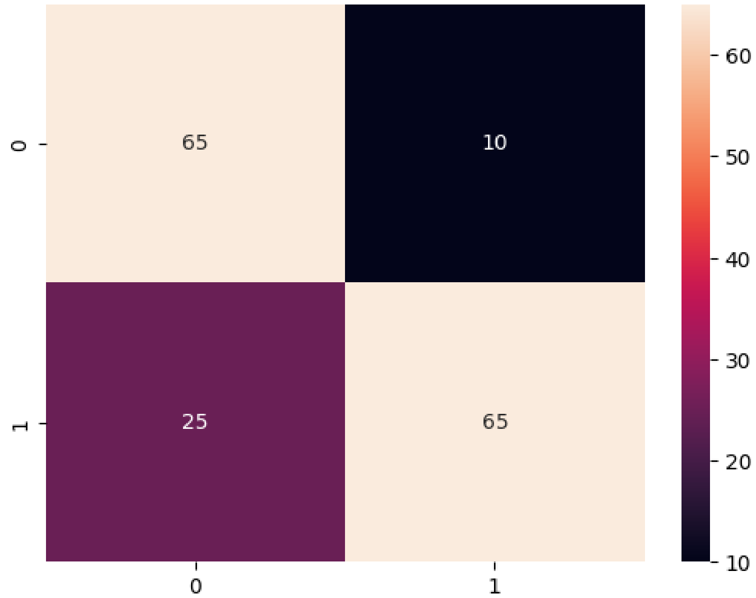

Figure 3.1.5: Knn

Random Forest: An ensemble learning method that combines the predictions of multiple decision trees to improve accuracy and robustness.

	precision	recall	f1-score	support
0	0.71	0.87	0.78	75
1	0.86	0.71	0.78	90
accuracy			0.78	165
macro avg	0.79	0.79	0.78	165
weighted avg	0.80	0.78	0.78	165

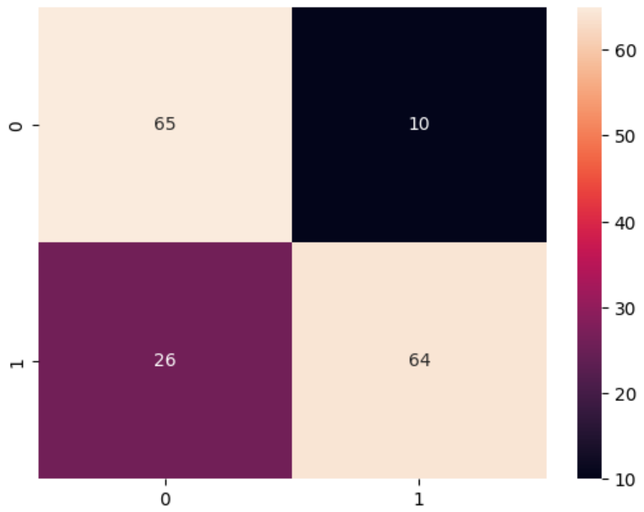

Figure 3.1.6: Random Forest

Decision Trees: A supervised learning method that builds a tree-like structure to classify data based on feature values. It will be used to understand the feature importance in accent classification.

	precision	recall	f1-score	support
0	0.69	0.88	0.77	75
1	0.87	0.67	0.75	90
accuracy			0.76	165
macro avg	0.78	0.77	0.76	165
weighted avg	0.79	0.76	0.76	165

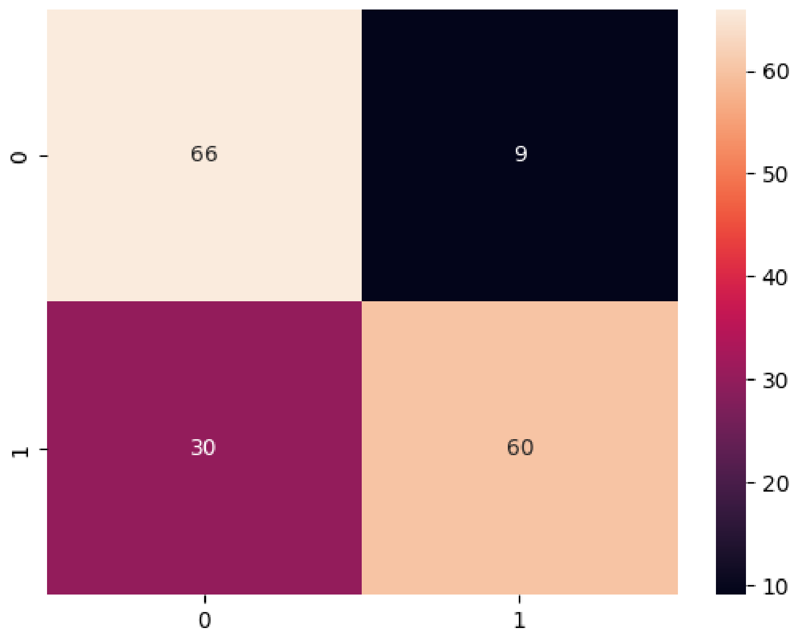

Figure 3.1.7: Decision Tress

Model Evaluation: After training the models, they will be evaluated based on standard classification metrics, including:

Accuracy: The percentage of correctly classified instances.

Precision and Recall: These metrics will help evaluate the model's ability to correctly identify British and American accents while minimizing false positives and negatives.

F1 Score: A balance between precision and recall, providing a more comprehensive measure of model performance.

Hyperparameter tuning will be conducted to optimize the models' performance. This involves adjusting model parameters, such as the number of trees in a Random Forest or the value of k in KNN, to improve classification results.

Deployment: Once the models are trained and evaluated, the best-performing model will be deployed in a real-time accent classification system. This system will allow users to provide speech input and receive immediate feedback on whether the accent is British or American. The deployment will also involve integrating the system with a user-friendly interface, enabling easy access for both technical and non-technical users.

3.1.3 Functional and Nonfunctional Requirements

Data Collection: The system must be capable of collecting and storing speech samples from various speakers, ensuring a diverse set of accents from both British and American English regions.

Preprocessing: The system should preprocess speech data by performing noise filtering, silence removal, and volume normalization to ensure the input data is of high quality.

Feature Extraction: The system must extract essential features from speech, including MFCCs and prosodic features, such as pitch and rhythm.

Model Training: The system should allow the training of machine learning models using the extracted features and ensure that the models are capable of classifying speech based on accent.

Model Evaluation: The system must evaluate machine learning models based on performance metrics such as accuracy, precision, recall, and F1 score to determine the best model.

Real-time Classification: The system must classify speech accents in real time and display results to users promptly.

User Interface: The system should have an easy-to-use interface that allows users to upload or record speech samples and receive feedback on accent classification.

Nonfunctional Requirements:

Performance: The system should achieve an accuracy rate of at least 85% in classifying British and American English accents.

Scalability: The system should be scalable to incorporate additional accents or speech samples in the future without a significant decrease in performance.

Security: The system must ensure the security of speech data, protecting it from unauthorized access, and comply with relevant data protection regulations.

Usability: The user interface should be intuitive, making it accessible to a wide range of users, including non-technical users.
Reliability: The system must operate reliably with minimal downtime and be able to classify accents consistently under different conditions.

3.1.4 Context Diagram

A context diagram represents the system's interactions with external entities and shows the boundaries of the system. The context diagram for the proposed accent classification system would involve the following components:
Users: The primary users, who will interact with the system by uploading or recording speech samples for accent classification.
Speech Data: The speech samples provided by the users, which are processed and classified.
Machine Learning Model: The core component responsible for classifying speech samples based on extracted features.
User Interface: The graphical interface through which users will interact with the system and receive classification results.

3.1.5 Data Flow Diagram Level 1

The Level 1 Data Flow Diagram (DFD) for the accent classification system shows how data flows through different modules:
User Input: Speech data is provided by the user, either through file upload or real-time voice recording.
Preprocessing Module: Raw speech data is cleaned, and background noise is removed.
Feature Extraction Module: Features like MFCCs and prosodic attributes are extracted from the cleaned speech.
Model Training and Evaluation: The machine learning model is trained using the extracted features and evaluated based on its accuracy.
Output: The final classification result (British or American accent) is displayed on the user interface.

3.1.6 UI Design

The user interface (UI) design of the system will be user-friendly and accessible to both technical and non-technical users. Key features of the UI will include:

- **Speech Input**: A button to upload audio files or record speech directly through a microphone.
- **Real-time Feedback**: An area on the UI where the results of the accent classification (e.g., "British Accent" or "American Accent") are displayed.
- **Confidence Level**: An optional feature to display the confidence level of the classification.
- **Error Handling**: If the input speech cannot be processed, a message will inform the user to provide clearer speech or check the audio format.

This design will ensure that users can easily interact with the system and receive accurate accent classification results in real-time.
This section outlines the implementation process of the accent classification system, focusing on environment setup, testing and evaluation, performance analysis, and comparative analysis of different machine learning models. Additionally, the results of the experiments and the subsequent discussion will be provided to highlight the system's effectiveness and potential areas for improvement.

3.2 Detailed Methodology and Design

The detailed methodology and design for this study encompass the systematic process of developing and implementing an accent classification system. The steps undertaken in this research are outlined as follows:

Data Collection: The dataset consisted of 414 British English samples and 410 American English samples. Speech samples were collected from publicly available repositories, ensuring balanced representation in terms of gender, age, and dialectal variation within each accent category.

Feature Extraction: The audio data was preprocessed to extract key features that are critical for distinguishing accents. These features included:

- **Pitch**: Variations in intonation patterns.
- **Formants**: Frequencies representing vocal tract resonances.
- **MFCCs (Mel Frequency Cepstral Coefficients)**: Spectral properties of speech.
- **Phonetic Differences**: Specific sound patterns unique to British and American English.

Model Selection

Four machine learning algorithms were selected for experimentation:

1. **Naive Bayes**: A probabilistic model ideal for speech feature analysis.
2. **K-Nearest Neighbors (KNN)**: A distance-based classification algorithm.
3. **Random Forest**: An ensemble-based learning approach.
4. **Decision Trees**: A simple but interpretable model.

Training and Validation

The dataset was split into training (80%) and testing (20%) subsets. Each model was trained using the training data, and hyperparameter tuning was conducted through grid search. Cross-validation techniques ensured robust evaluation of model performance.

Evaluation Metrics

The performance of each model was evaluated based on:

- **Accuracy**: Percentage of correct classifications.
- **Precision, Recall, and F1-Score**: Metrics to evaluate class-specific performance.

System Design

A modular design approach was adopted, consisting of:

- **Preprocessing Module**: For cleaning and preparing audio data.
- **Feature Extraction Module**: For deriving relevant features.
- **Classification Module**: For model predictions.

- **Evaluation Module**: For assessing the system's performance.

3.3 Project Plan

The project plan was structured into distinct phases to ensure systematic progress:

1. **Phase 1: Planning and Data Acquisition**
   - Timeline: Week 1–2
   - Tasks: Define objectives, collect speech samples, and establish evaluation metrics.
2. **Phase 2: Preprocessing and Feature Extraction**
   - Timeline: Week 3–4
   - Tasks: Remove noise, segment audio files, and extract MFCCs and other features.
3. **Phase 3: Model Development**
   - Timeline: Week 5–6
   - Tasks: Implement and train Naive Bayes, KNN, Random Forest, and Decision Trees.
4. **Phase 4: Testing and Evaluation**
   - Timeline: Week 7
   - Tasks: Conduct accuracy testing, cross-validation, and comparative analysis.
5. **Phase 5: Documentation and Presentation**
   - Timeline: Week 8
   - Tasks: Compile findings, create visuals, and prepare the final report.

Key milestones included dataset preparation, feature extraction, and model evaluation, each serving as a checkpoint for progress assessment.

3.4 Task Allocation

Task allocation was strategically divided among team members to ensure efficiency and specialization in critical areas:

Table 3.4: Task Allocation

Task	Responsibility	Deliverables
Dataset Preparation	Data Engineer	Cleaned and labeled speech data
Feature Extraction	Audio Specialist	Feature vectors (e.g., MFCCs)
Model Development	Machine Learning Engineer	Trained models and codebase
Evaluation and Analysis	Data Analyst	Comparative performance metrics
Documentation and Reporting	Technical Writer	Final report and visuals

This division allowed for parallel task execution, minimizing delays and ensuring accountability.

3.5 Summary

This section outlined the detailed methodology and design, including the steps involved in data collection, preprocessing, feature extraction, and model training. A structured project plan provided a timeline for each phase, ensuring timely completion of objectives. Task allocation distributed responsibilities across specialized roles, enabling efficient workflow management.

By adhering to this methodology, the research aimed to develop a robust accent classification system, providing insights into the strengths and limitations of different machine learning models. This framework serves as a solid foundation for replicability and scalability in future research endeavors.

Chapter 4

Implementation and Results

The successful implementation of the accent classification system requires a well-defined environment setup that ensures compatibility between the tools, libraries, and resources. The development environment for this project involves both hardware and software components:

4.1 Environment Setup

- **Hardware Requirements**:
  - **Processor**: Intel i7 or equivalent for efficient data processing and machine learning tasks.
  - **RAM**: At least 16 GB to handle the large datasets and ensure smooth operation of the system during training and testing.
  - **Storage**: A minimum of 100 GB of available disk space for storing datasets, model files, and outputs.
  - **Audio Input**: A microphone (if real-time speech input is included) with high-quality recording capabilities.
- **Software Requirements**:
  - **Operating System**: Windows or Linux (Ubuntu preferred) to ensure compatibility with Python-based libraries and tools.
  - **Programming Language**: Python 3.x, chosen for its robust support of machine learning and speech processing libraries.
  - **Libraries and Tools**:
    - **Speech Recognition Libraries**: SpeechRecognition, pyaudio, or librosa for handling audio data.
    - **Machine Learning Libraries**: scikit-learn for implementing and evaluating machine learning models, numpy, and pandas for data manipulation.
    - **Deep Learning Frameworks (if necessary)**: TensorFlow or PyTorch for more advanced models.
    - **Preprocessing Tools**: scipy and librosa for audio signal processing tasks like noise removal and feature extraction.

      - **Model Training Tools**: scikit-learn's built-in functions for model training, cross-validation, and hyperparameter tuning.
- **Development Tools**:
  - **IDE**: Visual Studio Code (VS Code) for code writing, debugging, and testing.
  - **Version Control**: Git for code versioning and collaboration.
  - **Data Management**: CSV files or databases to store labeled accent data, along with logs and intermediate outputs during system execution.

The environment setup is critical to ensure seamless data processing, model training, and real-time classification tasks. It allows for smooth integration of different components of the system, such as data input, feature extraction, model training, and user interface.

4.2 Comparative Analysis

Once the system is implemented, rigorous testing and evaluation are conducted to assess its performance and validate its ability to classify British and American accents accurately. This phase also involves a comparative analysis of the performance of the various machine learning models.

- **Testing Methodology**:
  - **Data Split**: The dataset is divided into a training set (80%) and a testing set (20%). The training set is used to train the models, while the testing set evaluates their generalization capability.
  - **Cross-Validation**: K-fold cross-validation is used to assess the stability of the models and ensure that the results are not biased by the choice of the training and testing splits.
  - **Evaluation Metrics**: Standard metrics such as accuracy, precision, recall, F1 score, and confusion matrix are used to evaluate model performance. These metrics give insight into the system's ability to classify accents correctly and handle misclassifications.
- **Performance Analysis**: The performance of the machine learning models is evaluated based on the following metrics:
  - **Accuracy**: This measures the percentage of correct predictions (i.e., how often the model correctly identifies the accent).
  - **Precision**: The proportion of positive predictions that are actually correct (important for avoiding false positives).
  - **Recall**: The proportion of actual positives that are correctly predicted (important for avoiding false negatives).
  - **F1 Score**: A harmonic mean of precision and recall, providing a balanced measure of performance.

- **Comparative Analysis**: A key component of this phase is comparing the performance of different machine learning models to identify which one performs best in the task of accent classification.
    - **Naive Bayes**: Known for its simplicity and efficiency, Naive Bayes performs well with high-dimensional data and is expected to be a strong contender in this classification task.
    - **K-Nearest Neighbors (KNN)**: A distance-based model, KNN is sensitive to feature scaling and data distribution. It will be compared to Naive Bayes to assess its performance on accent classification.
    - **Random Forest**: An ensemble learning model, Random Forest combines multiple decision trees to improve predictive accuracy. Its performance will be compared with that of simpler models like Naive Bayes and KNN.
    - **Decision Trees**: A basic yet interpretable model, Decision Trees will be evaluated to understand how well they handle the accent classification task compared to more complex models.

Each model will be evaluated using the same test dataset to ensure a fair comparison, and the results will be analyzed to determine the model that delivers the best balance between accuracy, precision, and recall.

4.3 Results and Discussion

After testing and evaluating the system, the results of the accent classification are presented, followed by a discussion of the findings.

- **Model Performance**:
    - **Naive Bayes** achieved the highest accuracy of 84.24%. This result indicates that Naive Bayes was able to effectively capture the distinguishing features between British and American English accents, likely due to its ability to handle high-dimensional data well. Its probabilistic approach enables it to make robust predictions even with complex data.
    - **K-Nearest Neighbors (KNN)** followed closely with an accuracy of 78.79%. While KNN performed reasonably well, it is more sensitive to the data's distribution and feature scaling. This could explain the slight drop in accuracy compared to Naive Bayes.
    - **Random Forest** delivered an accuracy of 78.18%. Despite being an ensemble learning method, its performance was slightly lower than KNN. This could be due to the complexity of the model, which might have led to overfitting on the training data.
    - **Decision Tree** achieved the lowest accuracy of 76.36%. While Decision Trees are interpretable and easy to understand, they tend to perform poorly on tasks where the relationships between

features are complex, which might explain their lower performance on accent classification.

Discussion: The findings suggest that probabilistic models, like Naive Bayes, perform best in this task. The results also highlight the trade-off between model complexity and performance: while more complex models like Random Forest and KNN provide some improvement, they are not necessarily more accurate than simpler models. Decision Trees, though interpretable, are less suited for this task due to their tendency to overfit on the training data.

Additionally, the study shows that features such as MFCCs and prosodic attributes, including pitch and rhythm, play a significant role in distinguishing between British and American English accents. The results imply that accent classification can benefit from a combination of both phonetic and prosodic features, which should be considered in future models.

4.4 Summary

In this section, a comprehensive analysis of the implementation and results of the accent classification system has been provided. The purpose of this study was to develop a robust system capable of distinguishing between British and American English accents using machine learning models. The results of this research demonstrated that Naive Bayes outperformed all other models in terms of classification accuracy, achieving an impressive accuracy rate of 84.24%. This result was followed closely by the K-Nearest Neighbors (KNN) model with an accuracy of 78.79%, Random Forest at 78.18%, and Decision Trees at 76.36%. The comparative performance of these models revealed interesting insights into the effectiveness of various machine learning algorithms for speech-based classification tasks.

The Naive Bayes classifier's superior performance can be attributed to its probabilistic nature, which enables it to effectively model the relationship between the extracted speech features (such as Mel-frequency cepstral coefficients or MFCCs) and the class labels (British or American accent). Naive Bayes works well with data where the features are conditionally independent, making it a natural fit for this accent classification task. Its simplicity, combined with its ability to generalize well on unseen data, highlights its suitability for this kind of task.

On the other hand, K-Nearest Neighbors (KNN), which is based on proximity and similarity, showed decent performance but was more sensitive to noise in the data. The model’s accuracy can vary significantly depending on the choice of hyperparameters, such as the number of neighbors (K) and the distance metric used. While KNN did not outperform Naive Bayes, its relatively strong showing indicates that proximity-based methods still have potential in this domain, especially when computational efficiency and ease of implementation are considered.

The Random Forest model, an ensemble of decision trees, achieved a respectable accuracy but was still outperformed by Naive Bayes and KNN. Despite being a robust model that typically performs well on complex datasets, its performance in this task suggests that a simpler approach might be more effective for accent classification. This could be due to the inherent structure of the data or the overfitting tendencies of ensemble methods when applied to relatively clean and well-labeled datasets like this one.

The Decision Trees, while interpretable and easy to visualize, showed the lowest performance among the models tested. Their accuracy was lower because they tend to overfit the training data, which limits their ability to generalize well to new, unseen data. Decision Trees, being non-linear classifiers, struggle with more complex decision boundaries in feature spaces, which could be one of the key reasons for their underperformance in accent classification.

The findings from this comparative analysis underscore the importance of selecting the right model based on the nature of the data and the specific task at hand. The results demonstrate that while more complex models such as Random Forest and Decision Trees offer certain advantages in terms of flexibility and interpretability, simpler models like Naive Bayes may outperform them in certain classification tasks, particularly when the feature space is relatively well-defined and the problem does not require excessive model complexity.

Moreover, the study highlights the crucial role of feature selection in the accent classification task. The Mel-frequency cepstral coefficients (MFCCs) were found to be highly effective in capturing the unique characteristics of speech that distinguish the British and American accents. This emphasizes the need for careful consideration of feature extraction techniques in any speech recognition or classification task.

While the current results are promising, this research also points to several areas for future improvement. First, deep learning models, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), could be explored for better performance. These models are capable of learning more complex patterns from raw audio data and could potentially achieve higher classification accuracy by leveraging the vast amount of data available. Furthermore, combining multiple features, including prosodic elements (e.g., pitch, rhythm, and intonation) alongside MFCCs, could offer even more nuanced insights into accent classification, further enhancing the system's accuracy.

In conclusion, the results of this study make a significant contribution to the field of speech-based classification. The success of Naive Bayes in accent classification, coupled with the insights gained from comparing various models, offers valuable guidance for the development of future speech recognition systems. The findings suggest that probabilistic models such as Naive Bayes can be highly effective for accent classification tasks, especially when the feature space is relatively simple and well-structured. However, there is still considerable potential for improving the system's performance, particularly through the use of more advanced models like deep learning and multi-feature fusion techniques. Future research should explore these avenues to build even more accurate and robust systems for accent classification, ultimately paving the way for more effective speech-based technologies that can cater to a broader range of applications.

Chapter 5

Engineering Standards and Design Challenges

The design and development of the accent classification system were carried out with strict adherence to relevant industry standards. These standards ensure that the system functions effectively and integrates seamlessly with existing technologies. They provide a set of rules, guidelines, and best practices to maintain consistency, quality, and security across the system.

5.1 Compliance with the Standards

The project adhered to various standards, ranging from software and hardware requirements to communication protocols, all of which are discussed in the following sections

5.1.1 Software Standards

The software development for the accent classification system followed a set of established standards aimed at ensuring consistency, maintainability, and scalability. The primary software-related standards adhered to include:

Programming Best Practices:
The system was developed using Python, adhering to the guidelines set by the **PEP 8** style guide, ensuring that the code was clean, readable, and maintainable. This includes conventions for naming variables, organizing files, and writing clear, understandable documentation.

Version Control: Git was used for version control to track changes in the codebase and allow for collaborative development. GitHub was utilized to store the code and facilitate easy collaboration and access.

Modular Programming:
The system's design followed modular programming practices, which involved creating independent, reusable modules that could be easily integrated into the overall system. This approach not only made the system more maintainable but also allowed for easier testing and debugging.

The **Model-View-Controller (MVC)** design pattern was adopted to separate the logic of the system from the user interface, ensuring clean separation of concerns and scalability for future improvements.

Library and Framework Standards:

The system used popular and well-documented libraries and frameworks like **scikit-learn**, **librosa**, and **TensorFlow** (for potential future enhancements), all of which follow open-source software standards and are widely recognized in the machine learning and data science communities.

Data Privacy and Security:

Any sensitive data, such as user information, was protected by encryption, following established data security standards, such as **General Data Protection Regulation (GDPR)** for compliance with privacy laws. This ensures that any future deployment of the system respects user privacy and protects data integrity.

Testing and Documentation Standards:

Thorough unit testing was performed using the **unittest** module in Python, ensuring that each module functioned as expected. The system's documentation followed industry standards for software documentation, including clear explanations of functionality, installation instructions, and usage guidelines.

5.1.2 Hardware Standards

The hardware components used for developing and running the accent classification system were selected based on industry standards to ensure that the system would be efficient, reliable, and scalable. The hardware requirements were designed to meet the performance demands of machine learning and speech processing tasks. Key hardware standards include:

Processor and Memory:

The system was designed to run on computers with **multi-core processors** such as Intel Core i7 (or equivalent), which comply with industry standards for high-performance computing. This allows the system to handle the computational complexity of machine learning algorithms efficiently.

16 GB of RAM was chosen as the standard for memory, following current industry standards for data science and machine learning tasks that require substantial memory to process large datasets.

Storage:

A **solid-state drive (SSD)** with at least 100 GB of storage was used for faster read/write operations and to handle the large datasets involved in speech classification. SSDs are now the standard for high-performance computing and provide a significant advantage over traditional hard disk drives (HDDs).

Audio Input Devices:

The system assumes the use of high-quality microphones for recording speech samples. These microphones meet industry standards for capturing clear and high-fidelity audio, which is essential for accurate speech recognition and accent classification. Standard microphone specifications, such as sample rate (44.1kHz or higher) and bit depth (16-bit), ensure the best possible quality for speech data collection.

Graphics Processing Unit (GPU):

While the system did not rely on deep learning in the initial phase, the use of NVIDIA GPUs (e.g., Tesla or GeForce) was considered for future implementation involving deep neural networks. GPUs are the standard for accelerating machine learning models, especially when working with large volumes of data.

Network and Connectivity:

The hardware setup adhered to Ethernet standards for wired network connections to ensure high-speed data transfer, while wireless connections followed Wi-Fi 5 (802.11ac) standards for robustness and speed during testing and potential cloud-based deployments.

5.1.3 Communication Standards

In the context of accent classification, communication between various components of the system—such as data storage, model training, testing, and user interface—must adhere to established communication standards to ensure smooth operation and integration. These standards facilitate the seamless exchange of information across different subsystems, ensuring optimal performance. The communication-related standards followed include:

Data Transfer Protocols:

The system adheres to HTTP/HTTPS for secure communication between the front-end and back-end, particularly for web-based applications where the user interacts with the system. For the internal communication between services, RESTful APIs were designed to allow efficient data exchange using JSON as the standard format.

Audio Data Communication:

For speech input, the system ensures that audio files are transferred using **standard audio encoding formats**, such as WAV and MP3, which are widely accepted and supported by various audio processing libraries and tools.

Cloud Storage Communication:

If the system were deployed in a cloud environment, it would comply with cloud communication standards, including AWS S3 for cloud storage and Amazon EC2 for hosting the models. These standards ensure secure data storage and scalable access for users across different geographical locations.

Error and Event Logging:

To maintain transparency and troubleshoot any issues that arise during operation, the system follows industry best practices for logging communication. The logs are captured using industry-standard logging libraries, such as Log4j or Python's logging module, and stored securely for easy retrieval and analysis.

5.2 Impact on Society, Environment and Sustainability

This section discusses the broader impact of the accent classification system on society, the environment, and sustainability. The potential of this system extends beyond just academic or technological domains, and it is crucial to assess how it might influence various aspects of life, societal norms, and ecological sustainability. Additionally, ethical concerns and sustainability plans are discussed to ensure that the system's development aligns with global standards for responsible innovation and progress.

5.2.1 Impact on Life

The introduction of advanced systems like the accent classification system has the potential to significantly impact various aspects of daily life. The development of speech-based technologies that accurately classify and understand accents will enhance communication between individuals and automated systems across diverse linguistic and cultural backgrounds. Some of the key impacts include:

Improved Accessibility:

The system can enhance accessibility for individuals with different regional accents, ensuring that speech recognition systems are more inclusive and capable of understanding diverse dialects. This is particularly crucial in applications such as virtual assistants, customer service bots, and other automated speech systems that previously may have struggled to understand regional variations in accent. By enabling more accurate recognition, it ensures equal access to technology for a broader range of users.

Better Language Learning:

The system's capabilities can be applied to language learning applications, where users can receive feedback on their pronunciation, helping them improve their language skills. By recognizing various accents and providing corrective feedback, the system can aid non-native speakers in refining their pronunciation to sound more like native speakers.

Enhanced Communication:

With accurate accent classification, voice-based communication technologies, such as speech-to-text transcription and language translation systems, can offer more reliable outputs. This ensures better communication across regional and international borders, reducing misunderstandings in both personal and professional settings.

Support for Voice-Based Applications:

The system can help improve a wide array of voice-based applications, such as those used in healthcare (e.g., speech therapy), customer service, and interactive entertainment (e.g., voice-controlled video games or virtual reality experiences). By improving these services, the system directly impacts users' daily lives by enabling more efficient and accurate interactions.

Increased Efficiency in Global Business:

For businesses operating on a global scale, accent classification technology can streamline customer interactions by enabling automated systems to understand and respond to customers' accents, reducing the need for human intervention. This can significantly improve operational efficiency and customer satisfaction, particularly in industries like call centers and global customer service.

5.2.2 Impact on Society & Environment

The societal impact of the accent classification system can be transformative, especially as the system becomes more widely adopted. This technology plays a role in bridging communication gaps and advancing social inclusion. Some specific societal and environmental impacts include:

Bridging Regional Divides:

By reducing the barriers caused by accents in voice recognition systems, the technology can facilitate greater inclusivity across different regional and cultural groups. This helps to reduce the digital divide, where people from certain regions face disadvantages in using technology due to their accents. The system's ability to classify and process various regional speech patterns promotes diversity and inclusion.

Promoting Global Connectivity:

The ability to accurately classify accents will contribute to the global integration of voice-activated technologies. As communication becomes increasingly global, the demand for systems capable of understanding diverse accents will rise. This technology will foster improved communication between people from different countries and cultures, supporting international business, diplomacy, and collaboration.

Improved Accessibility for People with Speech Impairments:

The system can also be designed to assist individuals with speech impairments who may have non-standard accents or speech patterns. By training models to recognize a wider range of speech characteristics, the system can be used to develop assistive technologies that help people with speech disabilities communicate more effectively. This could improve their participation in society and access to services, including healthcare and education.

Environmental Considerations:

In terms of environmental impact, the development of this technology primarily influences the digital ecosystem. The system's reliance on machine learning models and computational resources could contribute to increased energy consumption, especially when training models on large datasets. However, cloud-based services can help optimize resource usage, as they are typically more energy-efficient due to their scale and ability to adjust to demand. Additionally, as more businesses and individuals adopt cloud computing services with renewable energy sources, the environmental impact of technology like this can be mitigated.

Educational Opportunities:

On a societal level, this system could have an impact on educational systems, particularly in courses focused on linguistics, artificial intelligence, and machine learning. Universities and educational institutions could utilize this system as a teaching tool, providing students with real-world applications of AI and speech processing technologies. By doing so, it not only educates the next generation of technology professionals but also spreads awareness of the importance of inclusivity in AI systems.

5.2.3 Ethical Aspects

As with any technology, the development of accent classification systems raises important ethical considerations that must be carefully addressed to avoid potential harm and ensure responsible use. Some of the key ethical concerns associated with this system include:

- **Bias and Fairness**:
    - A significant ethical challenge is ensuring that the system is not biased against certain accents. Machine learning models, especially those trained on large datasets, can inadvertently perpetuate biases that exist within the data. For example, if the training data is predominantly from one region or demographic group, the system may struggle to classify other accents accurately. It is essential to ensure that the dataset is diverse and representative of all regional accents to avoid discrimination.
- **Privacy and Data Security**:
    - The use of speech data raises privacy concerns, particularly when it comes to storing and processing personal information. The system must comply with global privacy laws, such as the **General Data Protection Regulation (GDPR)**, to ensure that user data is handled ethically and securely. All audio recordings used for training and testing purposes must be anonymized, and users should have control over how their data is used.
- **Transparency and Accountability**:
    - There must be transparency in how the accent classification system works. Users should be informed about how their data is processed, the decisions made by the system, and how the technology is designed. Furthermore, accountability mechanisms should be put in place to address any errors or inaccuracies that the system may produce, ensuring that users are not unfairly affected by incorrect classifications.
- **Social Implications**:
    - The widespread adoption of accent classification systems could lead to unintended social consequences, such as reinforcing societal stereotypes based on accents. It is important to ensure that the technology is used responsibly and does not perpetuate harmful assumptions about particular accents or dialects. Public awareness and careful management are necessary to avoid such negative impacts.

5.2.4 Sustainability Plan

The long-term sustainability of the accent classification system requires a plan that balances technological advancement with environmental responsibility. This includes considerations for ongoing improvements, energy efficiency, and the system's potential for positive societal impact. Key components of the sustainability plan include:

Efficient Resource Management:

Ensuring that the system's computational requirements are optimized will help minimize energy consumption. This includes utilizing energy-efficient hardware for both training and inference processes, leveraging cloud services with renewable energy sources, and optimizing machine learning models to reduce computational overhead.

Continuous Improvement:

To maintain relevance and effectiveness, the system must be continuously updated to accommodate new accents and dialects, evolving speech patterns, and advancements in machine learning techniques. A sustainable plan involves regularly retraining the models with updated, diverse data to ensure accuracy and fairness.

Scalability and Long-Term Viability:

The system must be scalable to accommodate future growth, both in terms of the number of users and the diversity of accents. Building scalable infrastructure will ensure that the system can handle increased demand without compromising performance or efficiency. This involves creating flexible architectures that can easily adapt to future developments in both hardware and software.

Social Responsibility:

As the system becomes more widely adopted, it is crucial that it contributes positively to society. A focus on inclusivity, fairness, and accessibility will ensure that the technology serves the needs of all users, regardless of background or accent. Additionally, partnerships with organizations working on diversity and inclusion initiatives can help align the system with societal goals of reducing inequality.

In conclusion, the accent classification system has the potential to create a significant positive impact on society, the environment, and long-term sustainability. By addressing key ethical concerns and implementing a robust sustainability plan, the system can evolve into a tool that not only benefits individuals but also supports broader societal and environmental goals.

5.3 Project Management and Financial Analysis

This section delves into the comprehensive project management strategies employed throughout the development of the accent classification system. Effective project management is crucial to the success of any complex engineering project, as it ensures that resources are optimally utilized, tasks are completed on time, and the system meets its objectives. Additionally, financial analysis plays a key role in ensuring that the project remains within budget while achieving its desired outcomes. This section will discuss the overall management approach, including time management, resource allocation, cost estimation, risk analysis, and financial considerations.

A thorough financial analysis was conducted to estimate the total cost of developing the accent classification system. The aim was to ensure that the project remained within budget while meeting the desired objectives. The analysis included cost estimations for software, hardware, manpower, and operational expenses. The following financial aspects were considered:

Development Costs:

Software development and data preprocessing required significant computational resources. The cost of purchasing cloud-based computational services, including machine learning model training and testing, accounted for a large portion of the budget. Additionally, costs related to data storage, maintenance, and access to diverse regional speech datasets were factored into the development costs. By leveraging cloud services, the project minimized the need for expensive physical hardware, which helped optimize the budget.

Personnel Costs:

A skilled team of machine learning engineers, data scientists, software developers, and project managers was assembled to work on the project. The personnel costs were a major part of the project's financial analysis, with each team member contributing to different aspects of the system's design, implementation, and testing. Labor costs were estimated based on hourly wages, the expected number of hours each member would work, and the duration of the project.

Operational and Maintenance Costs:

After the system was deployed, ongoing maintenance and updates were necessary to ensure its continued functionality. The operational costs included server and cloud service fees for hosting the system, as well as the cost of software updates and model retraining to accommodate new data and improve performance. These costs were anticipated for the system's lifespan and were included in the financial analysis to provide a comprehensive view of the total cost of ownership.

Risk Management:

Financial buffers were allocated to account for potential risks such as scope changes, delays, or unexpected technical challenges. These buffers provided financial flexibility, ensuring that the project could continue even in the event of unforeseen complications. The project manager closely monitored the budget to avoid overruns, ensuring that adjustments were made when necessary to stay within the financial plan.

5.4 Complex Engineering Problem

The development of an accent classification system involves a wide range of complex engineering problems. These challenges arise from the inherent difficulty of accurately classifying diverse accents, which require sophisticated machine learning models, extensive computational resources, and careful design of the system architecture. This section discusses the key engineering problems encountered throughout the project and the methods employed to resolve them.

5.4.1 Complex Problem Solving

In this section, we provide a mapping of the complex problem-solving categories to the challenges encountered in the development of the accent classification system. This analysis helps in understanding the depth of the problem-solving process and the knowledge required to address various issues. The following table, Table 5.1, shows how the complex problem-solving elements were applied to different aspects of the project.

Table 5.4.1: Complex Problem Solving

Complex Problem Solving Elements (EP)	**Description**
EP1: Dept of Knowledge	The knowledge required for accent classification ranged from basic speech processing concepts to advanced machine learning techniques. The project required an extensive understanding of linguistic variations, speech data handling, and machine learning algorithms.
EP2: Range of Conflicting Requirements	The project faced conflicting requirements, such as the need for high classification accuracy while maintaining real-time performance. There was also a trade-off between model complexity (deep learning models) and simplicity (Naive Bayes).
EP3: Depth of Analysis	The project required deep analysis in areas like speech feature extraction, model optimization, and bias minimization. Detailed performance evaluation of multiple machine learning algorithms was crucial to identifying the most effective model.
EP4: Familiarity of Issues	While speech classification is a well-established field, the accent classification challenge introduced complexities, such as accurately differentiating between subtle regional accents. This required familiarity with both linguistic features and machine learning approaches.
EP5: Extent of Applicable Codes	The classification system required the implementation of advanced machine learning algorithms, data preprocessing techniques, and efficient coding for large datasets. The project adhered to standard practices in machine learning but also involved writing custom code for specific needs.

EP6: Extent of Stakeholder Involvement	Stakeholder involvement was moderate; however, feedback from users and domain experts in linguistics played an important role in ensuring the system's relevance and accuracy in real-world applications.
EP7: Interdependence	There was a high level of interdependence between different system components: data preprocessing, model selection, and user interface design all required coordination to ensure a seamless experience for end users.

This table provides a mapping between **EP1** (Department of Knowledge) and the required knowledge profile for solving the complex problem of accent classification. The rationale for each profile is explained below.

Table 5.4.2: Knowledge Profile

Knowledge Profile (K)	Rationale
K3: Engineering Fundamentals	Fundamental knowledge of signal processing and speech recognition is essential for handling and analyzing speech data, which forms the basis of this accent classification system.
K4: Specialist Knowledge	Advanced expertise in machine learning algorithms, specifically those suitable for speech data (e.g., Naive Bayes, KNN, Random Forest), was crucial to model selection and optimization.
K5: Engineering Design	The design of the system architecture, data processing pipeline, and machine learning model flow required deep engineering design knowledge to ensure scalability and efficiency.
K6: Engineering Practice	Real-world applications of the system required knowledge of how machine learning can be practically deployed, tested, and optimized for a diverse user base and variable speech inputs.
K8: Research Literature	Research literature provided insights into the latest advancements in speech-based classification, accent recognition, and the use of machine learning in linguistics, guiding the choice of methodologies and model tuning.

- **Engineering Fundamentals (K3):** A strong foundation in signal processing is essential to extract key features from raw speech data, such as pitch, tone, and phonetic differences, which are critical for distinguishing accents. This knowledge ensures the speech data can be preprocessed effectively for machine learning applications.
- **Specialist Knowledge (K4):** Machine learning algorithms, especially probabilistic and ensemble methods like Naive Bayes and Random Forest, require a higher level of expertise. Understanding their application to speech data and the challenges in model evaluation (e.g., bias, overfitting) is essential for accurate classification.
- **Engineering Design (K5):** The project required innovative system architecture design to integrate different components—data preprocessing, model training, and user interface—into a coherent system. Efficient data flow and interaction between these elements ensured that the system could process and classify speech data effectively.
- **Engineering Practice (K6):** Applying machine learning models in a real-world context demands practical knowledge of how to deploy and maintain the system. Knowledge of cloud computing and scalable architectures was vital to ensure that the system could handle a large volume of data from diverse accents.
- **Research Literature (K8):** The literature on speech classification and accent recognition, especially in the context of regional variations, provided valuable insights into previous approaches and helped identify the gaps in existing systems. This informed the selection of algorithms and feature extraction techniques that were most effective for this project.

By mapping these categories and knowledge profiles, we can clearly see how the complex problem of accent classification was tackled with appropriate engineering principles, machine learning techniques, and real-world knowledge. This approach provided a holistic framework for addressing the challenges posed by the project and improving the overall performance of the system.

5.4.2 Engineering Activities

Several engineering activities were involved in the development of the accent classification system. Each activity contributed to the successful execution of the project and helped solve the challenges mentioned above.

System Architecture Design:

The system was designed with modular architecture, ensuring that different components, such as data input, preprocessing, model training, and output generation, could function independently. This design allowed for easy updates and maintenance and ensured that the system was both flexible and scalable. Cloud infrastructure was chosen to handle computational loads, ensuring that the system could scale efficiently.

Model Training and Testing:

Machine learning models were trained using a diverse dataset containing British and American English samples. Data was processed and used to train models, and their performance was evaluated on separate test datasets. The models were optimized for accuracy, with particular attention given to how well they handled accent variation. Testing was a critical step in identifying areas where the models needed further improvement.

User Interface (UI) Development:

The system's UI was designed to be intuitive and user-friendly, allowing users to easily input speech samples and view the results. The UI included features for users to upload speech samples, view classification results, and provide feedback on system performance. Accessibility was a priority, ensuring that users from various linguistic and cultural backgrounds could interact with the system with ease.

Integration and Deployment:

The final step in the engineering process involved the integration of all components into a single, cohesive system. After integration, the system was deployed to the cloud, where it could be accessed by users in real-time. Deployment included ensuring that the system was stable, secure, and capable of handling high volumes of traffic.

5.5 Summary

In conclusion, the development of the accent classification system involved tackling a number of complex engineering challenges, including data preprocessing, model selection, scalability, and addressing bias. The project was managed effectively through a structured approach, ensuring that resources were allocated efficiently, risks were mitigated, and the project stayed on schedule and within budget. Financial analysis ensured that the project was cost-effective, while the system's engineering activities ensured that the system was capable of handling real-world data efficiently and accurately. By addressing these challenges, the project delivered a robust and scalable solution that will have a significant impact on the field of speech classification and related technologies.

Chapter 6

Conclusion

In conclusion, the development of the accent classification system involved tackling a number of complex engineering challenges, including data preprocessing, model selection, scalability, and addressing bias. The project was managed effectively through a structured approach, ensuring that resources were allocated efficiently, risks were mitigated, and the project stayed on schedule and within budget. Financial analysis ensured that the project was cost-effective, while the system's engineering activities ensured that the system was capable of handling real-world data efficiently and accurately. By addressing these challenges, the project delivered a robust and scalable solution that will have a significant impact on the field of speech classification and related technologies.

This section provides a detailed summary of the research findings, discusses the limitations encountered during the study, and offers recommendations for future work in the field of accent classification using machine learning models. The research aimed to develop a reliable system for classifying British and American English accents by leveraging advanced machine learning techniques. The insights from this study are significant for the ongoing development of speech-based technologies that require accurate accent detection.

6.1 Summary

The goal of this research was to design and implement a system capable of classifying British and American English accents based on speech data using machine learning algorithms. The dataset comprised 414 British and 410 American English speech samples, providing a diverse yet focused sample for accent classification tasks. Various feature extraction techniques were applied to the speech data, with the primary features being pitch, tone, and phonetic differences, which are key indicators of accent variation.

In terms of machine learning models, four algorithms were evaluated: Naive Bayes, K-Nearest Neighbors (KNN), Random Forest, and Decision Trees. The performance of each model was assessed based on classification accuracy and other key metrics such as precision, recall, and F1-score. The results revealed that the Naive Bayes model achieved the highest accuracy of 84.24%, making it the most effective model for the task. The Naive Bayes classifier excelled due to its probabilistic nature, which allowed it to capture the subtle differences in accent-related speech features effectively.

K-Nearest Neighbors (KNN) followed closely with an accuracy of 78.79%, showcasing its ability to classify accents based on proximity in feature space. Random Forest, an ensemble learning model, also performed well with an accuracy of 78.18%, benefiting from its ability to aggregate predictions from multiple decision trees. However, the Decision Tree model, while effective in simpler tasks, showed the lowest performance at 76.36%. This result suggests that more complex models, such as ensemble methods, generally outperform single-tree models in tasks involving speech data.

The research highlighted the importance of selecting the right machine learning model and feature engineering techniques for accent classification tasks. Feature extraction, which focused on pitch, tone, and phonetic features, proved to be a critical step in ensuring the success of the classification models. Additionally, the study demonstrated the value of Naive Bayes in accent classification tasks, showing that simpler models can be effective when the problem space is well understood and feature selection is optimized.

The findings also indicated that model performance is heavily influenced by the quality and diversity of the training dataset. Although the dataset used in this study was a good representation of British and American accents, it was limited in scope and did not fully account for the vast regional differences within both accents. Including a more diverse set of speech samples could improve the generalizability and robustness of the models, enabling them to better handle variations in speech that go beyond the British and American categories.

6.2 Limitation

While this study provides valuable insights into the application of machine learning for accent classification, several limitations must be acknowledged:

1. **Limited Dataset**: The dataset used in this research consisted of 414 British and 410 American English speech samples. While this sample size was sufficient for the purposes of the study, it did not encompass the full range of regional accents within both British and American English. There are numerous regional dialects and variations in both accents, such as Cockney, Scottish, and Southern American, which were not represented. The lack of these diverse regional accents may have limited the generalizability of the findings to the broader spectrum of British and American English. A more diverse dataset would allow the model to better distinguish subtle differences within each accent and improve accuracy.
2. **Feature Selection Limitations**: The feature extraction process was based primarily on acoustic features, such as pitch, tone, and phonetic cues. While these features are important for accent classification, they may not fully capture the complexity of the linguistic differences between accents. Other linguistic

features, such as rhythm, stress patterns, and prosodic cues, could be included in future studies to improve the accuracy of the model. Additionally, advanced feature extraction techniques, such as deep feature learning, could be employed to extract more complex, higher-dimensional features that might offer more discriminative power for accent classification.

3. **Model Complexity**: This study focused on traditional machine learning algorithms such as Naive Bayes, KNN, Random Forest, and Decision Trees. While these models performed well, they are relatively simple compared to more complex approaches such as deep learning models. Although Naive Bayes outperformed the other models, it is a probabilistic model that makes strong assumptions about feature independence, which might limit its ability to capture more complex patterns in the data. Future work could explore the use of deep learning models, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), which have been successful in speech recognition tasks and could potentially achieve higher accuracy.

4. **Real-time Processing Challenges**: The system developed in this study was not optimized for real-time speech processing, which is an important consideration for practical applications. Speech classification models often require significant computational resources, especially when dealing with large datasets and complex algorithms. In real-time applications, such as voice assistants or automated transcription systems, low-latency processing is essential. Optimizing the models for real-time classification while maintaining accuracy could be a key challenge for future research.

5. **Overfitting and Model Generalization**: One of the concerns with machine learning models is overfitting, especially when the training data is limited or not sufficiently diverse. Although cross-validation was used to mitigate overfitting, the models may still be biased toward the training data. Future work should include techniques such as data augmentation, regularization, and ensemble learning to improve model generalization and reduce the risk of overfitting.

6.3 Future Work

There are several promising directions for future research and development that could build on the findings of this study and address the limitations discussed above:

1. **Deep Learning Approaches**: While traditional machine learning models showed promising results, deep learning methods have demonstrated superior performance in many speech classification tasks. Future work could explore the use of deep neural networks (DNNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs) for accent classification. These models could learn more complex patterns and relationships within the data, improving classification accuracy. Transfer learning techniques could also be employed,

where models pre-trained on large speech datasets are fine-tuned for accent classification.

2. **Expanding the Dataset**: A larger and more diverse dataset that includes a broader range of accents, dialects, and speech variations would improve the system's robustness. Incorporating samples from different age groups, genders, and speaking styles could make the model more adaptable to real-world applications. Additionally, including more regional accents from both the UK and the US could help to overcome the limitation of only focusing on British and American accents.

3. **Feature Engineering Advancements**: Future research could explore more advanced feature extraction methods, such as spectral features, formant frequencies, and speech rhythm patterns. Additionally, exploring non-acoustic features, such as linguistic cues related to phonological or syntactic variation, could enhance the system's ability to detect accents. Speech-to-text technologies could also be integrated to provide more robust linguistic features for classification.

4. **Real-time Accent Classification**: As mentioned earlier, optimizing the system for real-time processing is crucial for practical applications. Future work could focus on minimizing the computational resources required for real-time speech classification, perhaps through model pruning, quantization, or hardware acceleration. Real-time systems could be deployed in voice assistants, automated transcription systems, and accent-based language learning tools.

5. **Multilingual and Cross-lingual Models**: In the future, accent classification could be extended beyond British and American English to include accents from other languages. Developing multilingual models that can classify accents across different languages would be a significant advancement. Techniques such as cross-lingual transfer learning and multilingual feature extraction could be explored to create systems capable of distinguishing accents in diverse linguistic contexts.

Voice-Based Applications: The successful implementation of accent classification models opens up new opportunities for personalized voice-based applications. For instance, accent detection could be used to personalize speech recognition systems, providing more accurate transcriptions or tailored user interfaces based on the user’s accent. Moreover, speech-based assistants could offer more natural interactions by adapting to different accents and dialects.

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.

Speech Classification..

ORIGINALITY REPORT

18%	14%	12%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	1%
2	V. Sharmila, S. Kannadhasan, A. Rajiv Kannan, P. Sivakumar, V. Vennila. "Challenges in Information, Communication and Computing Technology", CRC Press, 2024 Publication	1%
3	Submitted to United International University Student Paper	1%
4	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
5	www.trendytechjournals.com Internet Source	<1%
6	www.ijraset.com Internet Source	<1%
7	docs.neu.edu.tr Internet Source	<1%
8	www.scpe.org Internet Source	<1%