GENDER DETECTION FROM BENGALI VOICE USING MACHINE LEARNING APPROACH

BY

NEELA SARKER
ID: 201-15-14103
AND

JANNATUL FERDOUS
ID: 201-15-13788

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised By

MOHAMMAD MONIRUL ISLAM
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

MD ASSADUZZAMAN
Lecturer (Senior. Scale)
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled "Gender Detection From Bengali Voice Using Machine Learning Approach", submitted by Neela Sarker and Jannatul Ferdous 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 14-07-2024.

BOARD OF EXAMINERS

Md Taimur Ahad

Dr. Md. Taimur Ahad (MTA) **Board Chairman**
Associate Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

14.7.24

Mohammad Monirul Islam (MMI) **Internal Examiner 1**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Jakaria
14.07.2024

Md. Jakaria Zobair (MJZ) **Internal Examiner 2**
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

14.07.2024

Nazibur Rahman **External Examiner**
Technical Lead- Database Administrator
Telenor-Grameen Phone Account

DECLARATION

We hereby declare that this project has been done by us under the supervision of **the Mohammad Monirul Islam, Assistant Professor, Department of Computer Science and Engineering, Daffodil International University**. We also declare thatneither this project nor any part of this project has been submitted elsewhere for the awardof any degree or diploma.

Supervised by:

Mohammad Monirul Islam
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Md. Assaduzzaman
Lecturer (Senior. Scale)
Department of CSE
Daffodil International University

Submitted by:

Neela Sarker
(Neela Sarker)
ID: -201-15-14103
Department of CSE
Daffodil International University

Jannatul Ferdous
(Jannatul Ferdous)
ID: -201-15-13788
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

We are grateful and wish our profound indebtedness to **Mohammad Monirul Islam**, **Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

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

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

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

ABSTRACT

Gender identification presents a notable challenge in signal processing, with recent attention shifting towards vocal feature analysis over traditional image classification methods. Acknowledging that gender classification encompasses more than fundamental frequency and pitch, the research explores the importance of feature selection, akin to dimensionality reduction, crucial for identifying gender-specific traits. This study delves into the efficacy and significance of machine learning algorithms in addressing voice-based gender identification, examining the relationship between vocal fold thickness, wavelength, and pitch perception, particularly in distinguishing male and female voices. Leveraging machine learning techniques, the feasibility of gender identification from voice signals is demonstrated, employing methods such as Discrete Fourier Transform, Mel-spaced filter-bank, and log filter-bank energies to extract MFCC features. Gender identification from natural voice holds considerable implications, especially in practical applications where gender differentiation is vital. While conventional voice-to-text conversion may not require gender detection, real-world scenarios demand it, aligning with Natural Language Processing, a subset of artificial intelligence. The methodology entails a systematic workflow, encompassing input audio files, pre-processing, feature extraction, model training, and testing with separate datasets. The study yields a remarkable dataset of 2384 samples from more than 50 speakers, including 607 male,663 child,451 third gender, and 663 female voices. This research underscores machine learning's potential in addressing gender identification from voice, emphasizing its significance across various artificial intelligence applications. Our dataset comprises individuals aged 0 to 42+ from diverse locations, meticulously categorized into age groups and gender categories. Utilizing a smartphone and audio recording software, we collect a comprehensive database, strategically divided into training and testing sets. Notably, SVM and Random Forest emerge as top performers, achieving accuracies of 96.28% and 97.72%, respectively.

TABLE OF CONTENTS

CONTENTS	Page No
Board of examiners	i
Approval	ii
Declaration	iii
Acknowledgements	iv
Abstract	v
CHAPTER1: INTRODUCTION	**1-9**
1.1 Overview	1-2
1.2 Background and Present State	2-3
1.3 Problem Statement	3-4
1.4 Objectives	4-5
1.5 Scope and Limitations	5-6
1.6 Report Organization	7-8
1.7 Summary	8-9

CHAPTER 2: LITERATURE REVIEW	**10-17**
2.1 Overview	10
2.2 Related Works	11-13
2.3 Comparison between existing works	13-14
2.4 Open Issues	14-15
2.5 Summary	25-17
CHAPTER 3: METHODOLOGY/ REQUIREMENT ANALYSIS & DESIGN SPECIFICATION	**18-26**
3.1 Overview	18-19
3.2 Proposed Methodology/ System Design	19-23
3.3 Hardware/ Software Requirement	24-25
3.4 Project Management and Financial Analysis	25-26
3.5 Summary	26
CHAPTER 4: IMPLEMENTATION	**27-40**
4.1 Overview	27
4.2 Train Model/ Prototype Design	27-38
4.3 System Testing/ Model Evaluation	38-39
4.4 Summary	39-40

CHAPTER 5: RESULT AND ANALYSIS **41-45**

5.1 Overview 41

5.2 Experimental/ Simulation Result 41-42

5.3 Performance/ Comparative Analysis 43-44

5.4 Summary 44-45

CHAPTER 6: IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY **46-51**

6.1 Impact on Life 46-47

6.2 Impact on Society & Environment 47-48

6.3 Ethical Aspects 48-49

6.4 Sustainability Plan 49-50

6.5 Summary 50-51

CHAPTER 7: CONCLUSION AND FUTURE WORK **52-55**

7.1 Conclusions 52-53

7.2 Further Suggested Works 53

7.3 Limitations/ Conflict of Interests 54

REFERENCES

APPENDIX A

COURSE OUTCOMES, COMPLEX ENGINEERING PROBLEMS (EP) AND COMPLEX ENGINEERING ACTIVITIES (EA) ADDRESSING

APPENDIX B

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.2.1: Work Flow	19
Figure 3.2.2: Data Visualization Male and Female	20
Figure 3.2.3: Data Visualization 3rd Gender and Child	20
Figure 3.2.4: Mfcc Work Flow (source google)	21
Figure 3.2.5: Mfcc Equation (source google)	22
Figure 3.2.6: Mfcc Values	22
Figure 3.2.7: Workflow of Gender Recognition	23
Figure 4.2.1: Data count	28
Figure 4.2.2: Random Forest	33
Figure 4.2.3: Confusion Matrix of Random Forest	30
Figure 4.2.4: Decision Tree	31
Figure 4.2.5: Confusion Matrix of Decision Tree	32
Figure 4.2.6: Gradient Boosting	33
Figure 4.2.7: Confusion Matrix of Gradient Boosting	33
Figure 4.2.8: Logistic Regression	34
Figure 4.2.9: Confusion Matrix of Logistic Regression	35
Figure 4.2.10: KNeighbors	36
Figure 4.2.11: Confusion Matrix of KNeighbors	37
Figure 4.2.12: SVM	37
Figure 4.2.13: Confusion Matrix of SVM	38
Figure 5.2.1: Diagram of Result	42

LIST OF TABLES

TABLES	PAGE NO
TABLE 2.3.1: Comparative Analysis with Previous Work	13
TABLE 3.4.1: Estimated Cost for Machine Learning based Project	25
TABLE 4.2.1: Sample of Data	28
TABLE 4.2.2: Score of Random Forest	30
TABLE 4.2.3: Score of Decision Tree	31
TABLE 4.2.4: Score of Gradient Boosting	33
TABLE 4.2.5: Score of Logistic Regression	35
TABLE 4.2.6: Score of KNeighbors	36
TABLE 4.2.7: Score of SVM	38
TABLE 5.2.1: Result of Algorithms	42
TABLE 5.3.1: Discussion of Accuracy	44

LIST OF ACRONYMS

ML	Machine Learning
SVM	Support Vector Machine
MFCCs	Mel-frequency cepstral coefficients

CHAPTER 1

Introduction

1.1 Overview

Voice recognition stands out as a prominent topic within the realm of Natural Language Processing (NLP). Speech is the main form of human communication and is produced by complex biological processes involving several bodily organs. Computers have difficulty identifying gender distinctions in speech, but the human brain does it with ease. Gender recognition is essential for computers and social interactions. Automated gender categorization is essential to modern technology and is used in many different applications [1]. In contrast to several speech detection systems that depend on word sequencing, this study creates a voice recognition application that utilizes an individual's voice wave frequency. Most remarkably, this tool uses the Bangla language to determine an individual's gender on its own. The human hearing system has the amazing capacity to identify a person's gender based only on what they say. [4] By utilizing the complex interaction between sound levels and frequencies, our brain can recognize cues that are exclusive to one gender as they pass through the auditory pathway [3]. Machines, on the other hand, are incapable of doing this on their own and need human assistance. But now that machine learning techniques have been developed, these artificial systems can mimic gender identification in humans.

A thorough approach to feature extraction is necessary for the successful identification of relevant aspects from the human voice, which is critical to successful gender recognition. The extraction of these crucial features from the speech is directly linked to the efficacy of machine recognition, demonstrating the crucial role feature selection plays in obtaining high recognition rates [8]. Voice-based gender detection offers a wide range of real-world uses. One important use is crime detection, where the ability to identify a person's gender by voice is crucial for investigating crimes committed through voicemails or phone conversations.

Although identity theft is a common practice among criminals, national security agencies may use this technology to monitor populations efficiently and classify individuals based

on gender, which can help solve these types of crimes. Furthermore, voice-based gender identification is important for demographic research [6]. It is possible to mechanically extract a country's demographic or census data from human speech, which can provide useful statistics on things like gender, education level, and the presence of disabilities. This application helps to improve demographic insights and expedites the data-collecting procedure [9]. Voice-based gender recognition is a useful tool in the field of healthcare, especially for online or mobile healthcare systems. Taking into consideration gender-specific illnesses such as vocal fold pathologies like cysts, which may be skewed toward a certain gender, such as being discovered mostly in female patients, healthcare providers may utilize this tool to administer treatments more precisely.
Gender identification technology is especially relevant in the context of Bengali language, which has a rich linguistic legacy and is spoken by millions of people, mostly in Bangladesh and the Indian state of West Bengal. When compared to languages that are researched more extensively, there is very less study in the field of gender recognition from Bengali speech, despite the significance of this task.

1.2 Background and Present State

Gender detection from voice signals has been a subject of interest in the fields of speech processing and machine learning for several decades. This interest stems from various practical applications, such as enhancing user experience in voice-based systems, improving the performance of speaker identification systems, and enabling personalized services.
The process involves analyzing audio signals to determine the gender of the speaker, typically categorized as male or female. Traditional methods for gender detection relied on handcrafted features extracted from the speech signal, such as pitch, formant frequencies, and spectral features. These methods often required significant domain knowledge and manual feature engineering to achieve acceptable accuracy. With the advent of machine learning and deep learning, automated feature extraction and classification techniques have revolutionized this field. Bengali, being the seventh most spoken language in the world, presents unique challenges and opportunities for speech processing research. The phonetic and prosodic characteristics of Bengali are distinct from other languages, necessitating

tailored approaches for effective gender detection. While considerable research has been conducted on gender detection for widely spoken languages like English and Mandarin, there is a relative scarcity of work focusing on Bengali. The current state of gender detection from Bengali voice has seen notable advancements with the application of machine learning and deep learning techniques. Researchers have leveraged various machine learning algorithms, including Support Vector Machines (SVM), Decision Trees etc, to enhance the accuracy of gender classification.

One of the key trends in recent research is the use of deep learning models, which automatically learn and extract features from raw audio data, thus minimizing the need for manual feature extraction, have shown promising results in capturing the intricate patterns in speech that are indicative of gender. For instance, a typical approach might involve preprocessing the audio data to extract MFCCs, which are then fed into a ml model for classification. The use of MFCCs as input features has been prevalent due to their effectiveness in representing the power spectrum of the speech signal. Moreover, the development of large, annotated datasets of Bengali speech has facilitated more robust training and evaluation of these models. Researchers are increasingly adopting data augmentation techniques to address the issue of data scarcity and improve model generalization. Despite these advancements, challenges remain in achieving high accuracy, especially in noisy environments and across diverse speaking styles and dialects. Future research is likely to focus on improving the robustness of these models, incorporating more sophisticated neural architectures, and exploring transfer learning from models trained on other languages.

1.3 Problem Statement

In response to widespread issues that affect a variety of industries, Voice to Gender Identification must be adopted immediately. The problem of phone calls and voicemails being inherently anonymous is a significant barrier to crime detection, making it more difficult for law enforcement to solve crimes.

There is a need for a more precise and efficient technique of extracting demographic data, especially with regard to gender, as the traditional methods used in demographic investigations are inherently inefficient. Moreover, companies struggle with the drawbacks of inadequate digital marketing tactics, which result from the absence of a reliable way to determine the gender of the client, diminishing the effectiveness of business dealings. In the medical field, treatment accuracy suffers in the absence of a speech analysis technology specifically designed to identify illnesses particular to a given gender. Voice to Gender Identification is a holistic solution that tackles these issues methodically in order to improve marketing tactics, streamline demographic insights, improve criminal resolution, and raise the accuracy and efficacy of healthcare precision to new heights.

1.4 Objectives

The "Voice to Gender Recognition System" is essentially motivated by the goal of addressing practical issues in a variety of fields and promoting technological innovations, which will ultimately lead to more effective, safe, and customized applications in the areas of healthcare, demographic research, commercial transactions, and crime prevention. The development of a "Voice to Gender Recognition System" is driven by a multifaceted set of motivations, each stemming from the potential applications and societal implications of such technology. At its core, the primary objective is to create a sophisticated and reliable tool capable of accurately identifying an individual's gender based on their voice. This ambitious goal is underpinned by several key motivations:

Enhancing Security Protocols: The ability to accurately detect gender from voice recordings holds significant implications for enhancing security protocols, particularly in law enforcement and crime prevention. By leveraging advanced voice processing algorithms, authorities can improve their capabilities in identifying and apprehending suspects, thereby bolstering public safety and security.Streamlining Demographic Data Collection: The "Voice to Gender Recognition System" has the potential to streamline the process of gathering demographic data, particularly in census-taking and population studies. By automating the identification of gender from voice samples, researchers and government agencies can obtain more accurate and comprehensive data, facilitating evidence-based decision-making in areas such as healthcare, education, and urban planning. Improving Marketing and Business Plans: Accurate gender detection from voice recordings can inform marketing strategies and business plans by providing insights into specific target audiences. Companies can tailor their products and services more effectively to different gender demographics, leading to improved customer engagement and satisfaction.

Advancing Healthcare Precision: The application of voice-based gender recognition technology can contribute to advancing precision in healthcare by providing additional information about conditions that may vary between genders. From personalized treatment plans to early detection of gender-specific health risks, this technology has the potential to improve healthcare outcomes and patient experiences.

Pushing the Limits of Machine Learning and Artificial Intelligence: The development of the "Voice to Gender Recognition System" is motivated by a broader ambition to push the limits of machine learning and artificial intelligence (AI) applications. By tackling complex challenges in speech recognition and gender classification, researchers aim to contribute to technical advancements in the field, paving the way for more sophisticated and intelligent systems in the future.

Overall, the motivation behind the "Voice to Gender Recognition System" extends beyond mere technological innovation. It is driven by a desire to address practical issues in diverse fields, ranging from security and healthcare to marketing and business, ultimately aiming to create more effective, safe, and customized applications with far-reaching societal impact. This section elucidates the multifaceted motivations underlying the research endeavor, setting the stage for a deeper exploration of its objectives and outcomes.

1.5 Scope and Limitations

Voice-based gender detection systems hold immense potential across various sectors and are poised to witness widespread adoption due to their multifaceted applications. The significance of such systems is particularly pronounced in law enforcement and security contexts, where they play a pivotal role in emergency calls and forensic investigations. By leveraging sophisticated audio analysis techniques, these systems enhance the ability to identify and apprehend offenders, thereby bolstering public safety and security measures. Moreover, the integration of voice-based gender detection technology represents a transformative tool in demographic studies and census data collection efforts. By expediting the acquisition of gender-related data, these systems provide governments and research organizations with a more comprehensive and precise understanding of population dynamics. This facilitates informed decision-making and policy formulation in diverse domains, ranging from healthcare to urban planning.

Beyond law enforcement and demographic studies, voice-based gender detection systems offer numerous benefits across various sectors. In the realm of human resources, for instance, these systems contribute to fostering impartiality and fairness in the recruitment process by anonymizing early applicant evaluations. Similarly, in educational settings, the incorporation of gender-based data collection mechanisms enhance the objectivity and effectiveness of research projects, ultimately supporting educational institutions in promoting evidence-based practices and informed decision-making.

Furthermore, sectors such as entertainment, gaming, and smart home technology stand to benefit from the implementation of voice-based gender detection systems. In entertainment and gaming, for example, personalized experiences can be tailored to users based on their gender identity, enhancing user engagement and satisfaction. Similarly, smart home appliances can offer more customized interactions and services, catering to the preferences and needs of individuals based on their gender.

In addition to these applications, voice-based gender detection systems hold promise in public safety, social media, and beyond. Public safety agencies can leverage gender-specific data obtained during distress calls to improve emergency response procedures and enhance situational awareness. Social media platforms, on the other hand, can utilize this technology to deliver personalized content and recommendations to users, thereby enhancing user engagement and satisfaction.

Overall, the broad utility of voice-based gender recognition systems underscores their transformative potential across various domains and sectors. By enabling increased productivity, efficiency, and creativity, these systems have the capacity to drive innovation and progress, ultimately shaping a more inclusive and technologically advanced future.

1.6 Report Organization

The report is structured into 7 chapters, each focusing on specific aspects of the research study:

CHAPTER 1: INTRODUCTION

This chapter provides a comprehensive introduction to the research on gender detection from Bengali voice using machine learning. It covers the background and current state of the field, identifies the research problem, outlines the objectives, and defines the scope and limitations. Additionally, it includes an organization guide for the report and a summary of the key points.

CHAPTER 2: LITERATURE REVIEW

In this chapter, existing literature relevant to gender detection from voice is reviewed. It includes an overview of related works, compares different methodologies and findings, identifies open issues in the current research, and concludes with a summary of the literature review.

CHAPTER 3: METHODOLOGY/ REQUIREMENT ANALYSIS & DESIGN SPECIFICATION

This chapter details the research methodology and design specifications. It presents the proposed methodology and system design, lists the hardware and software requirements, discusses project management strategies and financial analysis, and provides a summary of the methodology and design.

CHAPTER 4: IMPLEMENTATION

The implementation chapter describes the process of training the model and developing the prototype. It includes an overview of the implementation steps, details on model training and prototype design, and discusses system testing and model evaluation. The chapter concludes with a summary of the implementation.

CHAPTER 5: RESULT AND ANALYSIS

This chapter presents the experimental and simulation results of the study. It includes an overview, detailed results, performance and comparative analysis of different machine learning algorithms, and a summary of the findings.

CHAPTER 6: IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY

The societal, environmental, and ethical impacts of the research are discussed in this chapter. It covers the implications for individual lives, broader societal and environmental impacts, ethical considerations, and outlines a sustainability plan. The chapter concludes with a summary of these impacts.

CHAPTER 7: CONCLUSION AND FUTURE WORK

The final chapter summarizes the overall findings and contributions of the research. It suggests potential areas for future research, discusses the limitations of the study, and addresses any potential conflicts of interest.

Furthermore, the translation of research findings into practical applications represents a pivotal step towards realizing the societal impact of gender detection technology. Browser extensions, mobile applications, or voice-enabled devices equipped with real-time gender detection capabilities offer tangible solutions for users to interact with technology in intuitive and personalized ways. These applications are designed to empower users while respecting their privacy and cultural identity, thereby promoting inclusivity and user acceptance in diverse cultural contexts. Ultimately, the culmination of these efforts contributes to the broader field of gender detection technology by advancing our understanding of gender representation and identity in Bengali voice. By pioneering inclusive and ethically grounded solutions tailored to the Bengali-speaking community and beyond, this research fosters social progress and equity, paving the way for a more inclusive and technologically empowered social.

1.7 Summary

The significance of speech recognition is discussed in the introduction, with concentrate on the significance that it is to natural language processing. Comparing with normal speech detection systems, the study presents a novel voice recognition application that makes use of voice wave frequency, specifically in the Bangla language. The study examines how difficult it is for humans to distinguish gender from speech, comparing this to the limits of machines and the most recent developments made possible by machine learning. The potential social impact of voice-based gender recognition is highlighted by discussing its useful uses in marketing, healthcare, demographic research, and crime prevention. The problem statement highlights the need for a comprehensive Voice to Gender Identification solution by identifying the difficulties in criminal detection, ineffective demographic data collecting, inadequacies in digital marketing, and healthcare accuracy.

The goals and motivation center on developing a dependable instrument to deal with these issues and advancing technology. The project scope highlights the system's disruptive potential by envisioning broad applications across a range of industries, including social media and law enforcement. In summary, the project will be a multidimensional success, with improved marketing tactics, simplified demographic insights, enhanced criminal resolution, and advanced healthcare accuracy thanks to a smart Voice to Gender Recognition System.

CHAPTER 2

Literature Review

2.1 Overview

In our primary endeavor, we aim to explore the intricate landscape of human-based Bangla audio voices, specifically targeting gender identification within audio files. The potential societal impact of gender detection from human voices is far-reaching, holding the promise of transformative changes across diverse life domains. Our central focus revolves around the development of a specialized algorithm, propelled by the training of raw, physically collected data to elevate precision levels in gender identification. The research dataset undergoes a meticulous division into distinct training and testing groups, guided by predefined percentages to ensure robust evaluation. Employing a well-structured training and validation checking protocol, the training subset is further partitioned into additional percentage groups, ensuring a comprehensive examination of the algorithm's efficacy. This research stands at the critical intersection of technology and human communication, with the overarching goal of making substantial contributions to the advancement of gender recognition systems. The potential applications of our work span diverse societal contexts, reflecting a dedicated commitment to leveraging technology for positive societal impact and fostering innovation within the evolving field of voice-based gender detection.

2.2 Related Works

Decades of research in speech processing, a subset of digital signal processing, led to constructing a gender detection system. Extracting first formant and pitch using linear predictive analysis revealed higher values in female voices. While traditional K-nearest neighbor classification lacked flexibility, modern deep learning algorithms, as demonstrated in recent studies, overcome these limitations.[4]. A large dataset of audio files is used for gender detection, employing models like Support Vector Machine,Decision Trees, Gradient Tree Boosting, and Random Forest.

The process involves converting voice files to a machine-readable format, preprocessing to eliminate external noises, and extracting discriminative acoustic features. The machine is trained with these features using various algorithms, and accuracy is calculated to determine the best- performing algorithm for real-world gender detection [24]. A gender recognition system with SVM and PCA excels on a balanced dataset, reducing search spaces and response delays. Utilizing acoustic parameters, it achieves a remarkable 98.42% accuracy, emphasizing precision and recall. Tailored for efficient word identification, it accommodates short audio signals [5]. In paper [10] introduces a gender identification method utilizing SVM, Pitch, and MFCCs. Voiced frames are extracted using STE and ZCR, and experiments on a multilingual database demonstrate a 99.5% accuracy, surpassing previous claims. The proposed model enhances accuracy by incorporating pitch and MFCCs without separating voiced and unvoiced frames. In a study [11], an efficient time-domain gender detection algorithm is proposed, utilizing Autocorrelation (ACF) for pitch detection and K-Means as a classifier.

The algorithm introduces effective frame selection techniques to enhance both pitch detection and classifier efficiency. The method was tested on a forty-speaker speech database, demonstrating improved correctness in gender classification after implementing the proposed algorithm. The following study [19] presents a gender detection system utilizing acoustic parameters and employing four models—CART, XGBoost, SVM, and Random Forest.

Khan et al. [21] propose a privacy-preserving approach to gender recognition from speech using Federated Learning techniques. By decentralizing model training across multiple devices while keeping data localized, their method ensures user privacy while still achieving high accuracy in gender classification tasks, addressing concerns related to data privacy and security in voice-based recognition systems.

An ensemble approach enhances overall accuracy, positioning the system as a foundational element for various software applications, capable of extracting acoustic parameters and identifying the speaker's gender effectively. Voice Onset Time (VOT) values of stop consonants in a language can also provide clues about the gender of the speaker when used to identify that speaker [30]. In order to identify gender indicators, the VOT values of six stop consonants in the Turkish language were examined using classification methods such as SVM, KNN, and logistic regression.

Gupta et al. conduct a comprehensive comparative study of various machine learning algorithms for gender recognition from voice, including Decision Trees, Random Forests, SVMs, and KNN. Their findings shed light on the relative strengths and weaknesses of different algorithms in handling gender classification tasks across diverse datasets. [28] In this research, Li et al. investigate the application of Support Vector Machines (SVMs) for voice-based gender identification. By extracting acoustic features from speech signals and training SVM classifiers, the authors achieve high accuracy in gender classification tasks, highlighting the potential of SVMs in this domain.

These related works provide valuable insights into the application of machine learning and deep learning techniques for gender detection from Bengali voice. By building upon existing research findings and methodologies, our study contributes to the advancement of gender detection technology in the Bengali language context, paving the way for more accurate and culturally relevant solutions.

2.3 Comparison between existing works

Various studies in gender detection through speech analysis employ diverse techniques and models. The first work delves into traditional linear predictive analysis, highlighting the identification of higher pitch and formant values in female voices and the transition to more flexible deep learning algorithms. Another study employs a comprehensive dataset and multiple models (Support Vector Machine, Decision Trees, Gradient Tree Boosting, and Random Forest), emphasizing preprocessing for noise elimination and algorithm comparison for accuracy. A separate work introduces an SVM-based method with pitch and MFCCs, surpassing previous accuracy claims. Another focuses on an efficient time-domain algorithm utilizing Autocorrelation and K-Means for enhanced gender classification. Additionally, a gender recognition system with SVM and PCA excels with high accuracy and efficiency. Lastly, a study presents a foundational gender detection system using acoustic parameters and fourmodels, emphasizing the application's potential across various software, while another study explores Voice Onset Time values for gender identification using classification methods.

Table 2.3.1: Comparative analysis with previous work

No	Author	Algorithm	Result
01	Pramit et al (2018)	Random Forest	94.78%
02	Sharma et al (2020)	Support Vector Machine	92.47%
03	Gupta et al (2016)	Support Vector Machine	95.3%
04	Kumari et al (2015)	K-means clustering	90.70%
05	Kushwah et al (2019)	Support Vector Machine	89%
06	Nair et al (2019)	Gradient Boosting	93.70%
07	Korkmaz et al (2018)	Support Vector Machine	90%
08	Dutta et al (2013)	Logistic Regression	85.12%
09	Eckert et al (2022)	Decision Tree	90.70%
10	Vijayan et al (2018)	Cat Boost	89.59%
11	Weston (2016)	Support Vector Machine	94.69%

12	Pauly (2023)	Random Forest	92.83%
13	Mavadati (2019)	Decision Tree	87.26%

Overall, these studies contribute diverse insights into gender detection through speech analysis, ranging from traditional methods to sophisticated machine learningapproaches.

2.4 Open Issues

This paper employs four algorithms, namely Random Forest, Decision Tree, Logistic Regression, and KNeighbors algorithm, to precisely conduct our work. The raw data used in this study was collected in audio format and subsequently converted into mp3 format. The data was segmented into four categories: Child, Male, Female, and 3rd gender. The age range for the Child category was set between 0 to 5 years, while for the remaining categories, the age range was 42+. This work's materials and methods section comprises four parts [31]. The initial part provides details about the selected syllables. The second part addresses the participating speakers in this study. The third part outlines the procedure for collecting utterances. The final part references the method used for extracting VOT from syllables In the paper [6], About 3168 voice samples, initially .AIF files, were preprocessed using the "specan" function to extract 22 acoustic features. The resulting CSV file, with 3168 rows and 21 columns, enabled predicting output labels based on 20 features. Looking ahead, our study lays the foundation for future research endeavors aimed at enhancing the robustness and generalizability of gender detection algorithms in the Bengali language context. Addressing the ethical complexities and societal implications of gender detection technology will remain paramount, necessitating ongoing interdisciplinary collaboration, stakeholder engagement, and ethical reflexivity. By fostering a culture of inclusivity, equity, and responsible innovation, we can harness the transformative potential of machine learning in ways that empower individuals, foster cultural understanding, and promote social progress. The data was split into Train, Validation, and Test sets, with a selected ratio for optimal model results. While K-fold cross-validation could be considered, normal splitting sufficeddue to ample data availability. Extracting pitch values from voiced frames in speech signalsis crucial for accuracy, as voiced frames yield more precise results [11]. Utilizing ZCR andSTE, voiced frames are initially extracted from the speech signal, and pitch values are derived through cepstral analysis. However, pitch alone is inadequate for gender identification due to overlap in some cases. This paper employs pitch values and MFCC coefficients as a feature vector for training and testing the SVM classifier. The entire feature-set[16] underwent application of KNN and SVM. However, this approach might have led to an excessively flexible fit. While these models exhibited high performance, they couldn't uncover individual relationships. Despite the performance of previous models, utilizing a minimized set of features can still yield reliable results. This not only simplifies the model but also aids in understanding the significance of the remaining features in determining a person's gender.

2.5 Summary

One of the primary task in our research revolves around the collection and annotation of Bengali voice data. Bengali is a linguistically diverse language, encompassing various dialects and regional accents, which poses difficulties in assembling a comprehensive and representative dataset. Additionally, accurately annotating voice samples with gender labels can be subjective and prone to biases, requiring careful consideration of sociocultural norms and individual identities.

Another challenge is the presence of data imbalance and bias within the collected dataset. Gender distribution disparities, where certain gender categories are overrepresented or underrepresented, can skew the performance of machine learning algorithms, leading to biased predictions and inaccurate results. Addressing data imbalance and bias requires sophisticated data preprocessing techniques and algorithmic adjustments to ensure fair and equitable gender detection outcomes.

Extracting meaningful features from Bengali voice data poses a significant challenge due to the complex and nuanced nature of speech signals. While Mel-Frequency Cepstral Coefficients (MFCCs) serve as a widely used feature representation in speech processing tasks, capturing the rich linguistic and phonetic characteristics of Bengali speech requires more nuanced feature extraction techniques tailored to the language's specific acoustic properties.

Choosing appropriate machine learning algorithms and optimizing their parameters pose challenges in our research. While traditional algorithms such as Random Forest, Decision Tree, Logistic Regression, and K-Nearest Neighbors (KNN) exhibit promising performance, the advent of deep learning methodologies introduces new complexities and trade-offs. Selecting the most suitable algorithmic approach and fine-tuning its parameters to achieve optimal gender detection accuracy require extensive experimentation and computational resources.

Ethical and sociocultural considerations present ongoing challenges in our research. Ensuring privacy protection, informed consent, and cultural sensitivity in data collection and analysis is essential to uphold ethical research practices and respect the rights and dignity of participants. Moreover, navigating the complex sociocultural landscape of gender identity and expression in Bengali-speaking communities requires careful attention to avoid perpetuating stereotypes or biases in gender detection systems.

Finally, achieving robust generalization and real-world deployment of gender detection models present challenges in translating research findings into practical applications. Ensuring that trained models perform accurately across diverse linguistic contexts, environmental conditions, and user demographics necessitates rigorous validation and testing procedures. Moreover, developing user-friendly and culturally relevant applications that integrate seamlessly into everyday contexts requires interdisciplinary collaboration and user-centric design principles.

Addressing these challenges requires concerted efforts from researchers, technologists, and stakeholders to advance the field of gender detection from Bengali voice using machine learning approaches while ensuring ethical integrity, algorithmic fairness, and societal relevance. By overcoming these obstacles, we can unlock the transformative potential of technology to empower individuals, promote inclusivity, and foster social progress.by utilizing many models and a large dataset. Previous accuracy claims are exceeded by an SVM-based approach using pitch and MFCCs, and gender categorization is enhanced by an effective time- domain technique that makes use of this phenomenon and K-Means. Furthermore, a gender recognition system that combines PCA and SVM achieves excellent efficiency and accuracy. Another research highlights possible uses and proposes a basic gender detection system employing four models and auditory characteristics. Finally, voice onset time values are investigated in relation to gender identity. All things considered, these researches offer a variety of perspectives on gender identification by speech analysis, ranging from conventional to sophisticated machine learning techniques.

CHAPTER 3

Methodology Analysis & Design Specification

3.1 Overview

In the endeavor to uncover "Gender Detection from Bangla Voice Using a Machine Learning Approach," this study meticulously adheres to a systematic methodology. Commencing with the acquisition of a comprehensive Bangla voice dataset, the exploration spans both physical and digital channels to ensure a rich and diverse source of data. Subsequent preprocessing steps meticulously address the conversion of voice files into machine-readable formats, concurrently eliminating extraneous noise for heightened data quality. The distribution of samples across these categories, with 663 samples labeled as Child, 607 as Male, 663 as Female, and 451 as Third Gender, underscores the dataset's diversity and its relevance to real-world applications. To achieve gender detection, various machine learning algorithms are employed, including Random Forest, Decision Tree, Logistic Regression, SVM. Gradient Boosting and KNN. These algorithms are trained on the dataset, with MFCCs extracted from the voice samples serving as input features. The performance of each algorithm is evaluated based on metrics such as accuracy, precision, recall, and F1-score, providing insights into their effectiveness in distinguishing between different gender categories. Notably, the Random Forest algorithm achieves the highest, showcasing its robustness and suitability for gender detection from Bengali voice data.

The core of the methodology centers on feature extraction, where the focus is finely tuned to recognize pertinent acoustic parameters pivotal for gender identification. These parameters encompass the intricacies of Bangla voice patterns, including pitch, formant values, and Mel Frequency Cepstral Coefficients (MFCCs). The algorithmic groundwork is laid through a scrupulous selection process, integrating machine learning models like SVM, Decision Trees, and others fine-tuned to the nuances of Bangla vocal characteristics. Strategic division of the dataset into training and testing sets serves as the cornerstone for facilitating a comprehensive evaluation of the model's performance. Ethical considerations take precedence, ensuring strict adherence to ethical standards across data collection, usage, and interpretation, underscoring the study's commitment to a responsible and principled research approach.

3.2 Proposed Methodology

Throughout this research, handling raw audio data, initially in audible formats, posed distinct challenges. However, limitations in accessing specific audio formation systems, essential for further analysis, mandated a compulsory conversion of all files into a universally supported audible format, namely mp3. This adaptive measure proved crucial in overcoming accessibility constraints, ensuring the seamless continuity of the research endeavor.

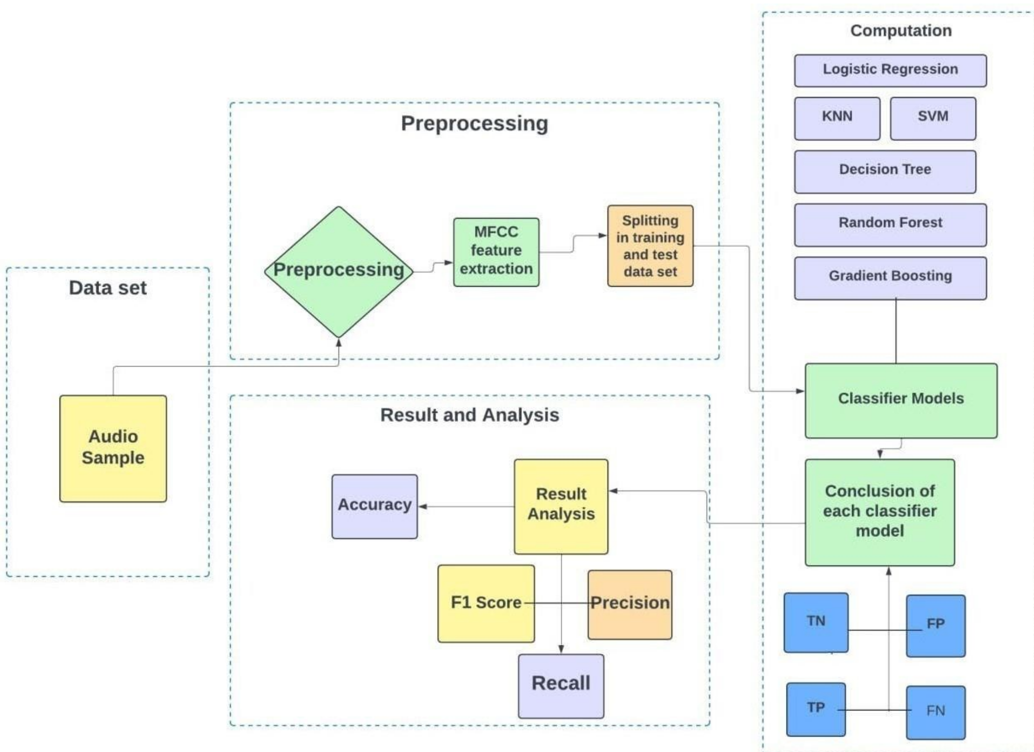

Figure 3.2.1: Work Flow

The conversion process played a pivotal role in making the data universally compatible for subsequent analysis and modeling, providing a foundation for robust and comprehensive research outcomes. To streamline data comprehension and organization, a meticulously executed data annotation or labeling process was set in motion. This involved the creation of distinct folders, each dedicated to a particular gender category—male, female, child, and third gender. This systematic structuring ensured an organized and easily identifiable framework for the dataset, laying the groundwork for subsequent analytical processes. The subsequent phase witnessed the transformation of audible data into numeric form.

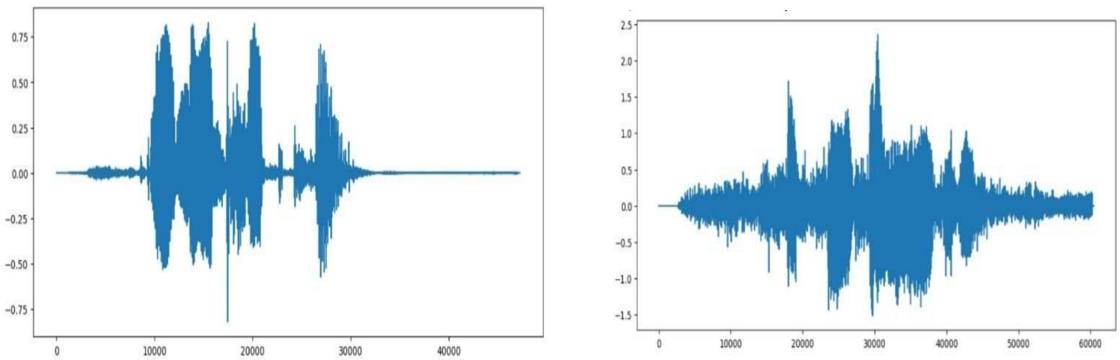

Figure 3.2.2: Data Visualization Male and Female

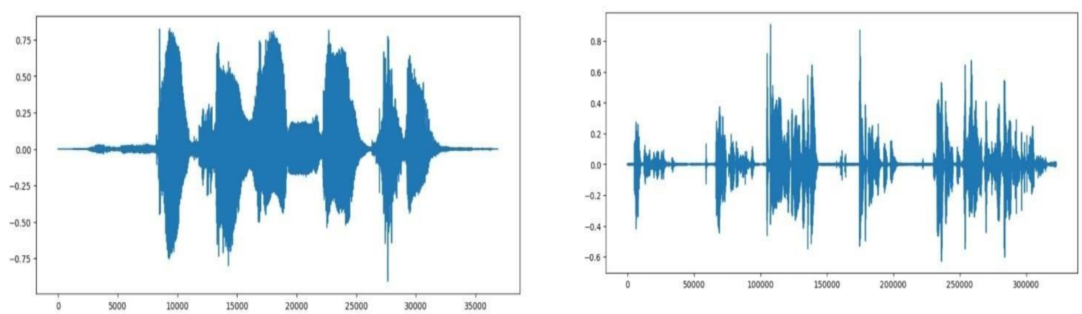

Figure 3.2.3: Data Visualization of 3rd Gender and Child

This pivotal step, referred to as feature extraction, demanded a nuanced approach. It involved the individual scrutiny of annotated data files corresponding to each specific gender category. Each file underwent a comprehensive transformation, translating audible nuances into numerical representations, thereby enabling the data to be processed and analyzed with precision.

There are many processes in the extraction of the MFCC: Pre-emphasis: To enhance high-frequency components, the input voice signal is pre-emphasized. Framing: The pre-emphasized signal is broken up into brief segments. Windowing: To minimize spectral leakage, a window function is multiplied by each frame. FFT: To acquire the power spectrum, the Fourier transform is done to each framed signal. Mel-filterbank: Filterbank energies are obtained by passing the power spectrum through a series of Mel-frequency filters. Logarithm: To get close to the logarithm of the human hearing response, the logarithm of the filterbank energies is calculated. Using the discrete cosine transform (DCT), the logarithmic filterbank energies are transformed to decorrelate the features. Feature Vector: The MFCC feature vector is composed of the lower-order DCT coefficients.

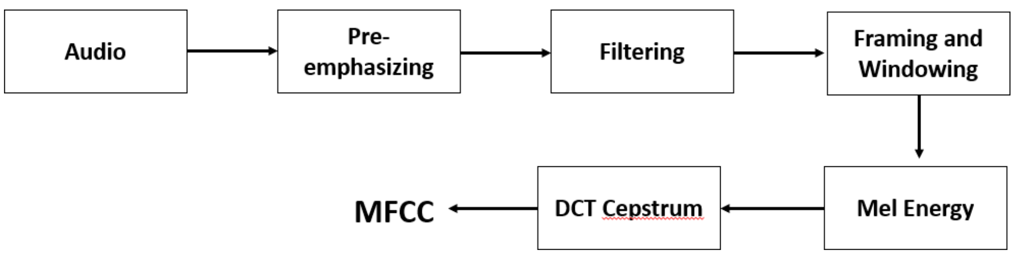

Figure 3.2.4: Mfcc Work Flow

Let $x(t)$ x(t) be the input audio signal, $w(t)$ w(t) be the window function, $X(\omega)$ X(ω) be the Fourier Transform of the windowed signal, $H_m(\omega)$ H m (ω) be the Mel filterbank,

C_m C m be the cepstral coefficients, and $DCT()$ DCT() represent the Discrete Cosine Transform.

$$x_n(t) = x(t) \cdot w(t - nT)$$
$$X_n(\omega) = \mathcal{F}(x_n(t))$$
$$E_m = \sum_{n=0}^{N-1} |X_n(\omega)|^2 \cdot H_m(\omega)$$
$$C_m = DCT(\log(E_m))$$

Figure 3.2.5: Mfcc Equation.

mfcc1	mfcc2	mfcc3	mfcc4	...	mfcc12	mfcc13	mfcc14	mfcc15	mfcc16	mfcc17	mfcc18	mfcc19	mfcc20
-387.480835	105.846504	-6.318912	24.977522	...	-6.445422	-7.009394	0.900065	-5.536519	-7.087941	-1.713805	-5.958529	-5.460635	-4.129093
-393.786194	123.849815	-7.053511	28.355192	...	-6.157703	-5.728826	-2.963459	-5.852234	-4.966861	-3.768545	-6.360138	-6.181791	-7.094828
-380.721466	93.914711	-11.127810	22.912466	...	-3.597420	-5.382451	-2.216980	-1.841336	-5.596298	-1.491200	-4.575242	-4.782634	-3.359289
-380.056213	111.787079	-17.421543	35.162041	...	-5.270954	-6.609304	-5.116140	-5.538610	-2.872726	-4.079984	-9.488663	-3.507008	-5.590202
-411.116486	88.305618	1.051208	24.495373	...	-7.752732	-6.472996	-1.118252	-7.120439	-5.633917	-3.657744	-5.274695	-4.157248	-2.881569
...	...	...	...	...	...	...	...	...	...	...	...	...	...
-252.923264	147.694275	-15.206455	6.682262	...	-2.453038	-7.266009	-0.511708	-4.190579	-4.420167	-3.859631	-5.829065	-1.951795	-5.161557
-275.773102	149.538269	-7.569076	-3.637012	...	-7.104940	-7.399678	0.892414	-7.165701	-6.074296	-2.691834	-6.231872	-7.936574	-3.290691
-233.282288	164.431854	-23.704670	2.624442	...	1.046293	-11.268488	-0.909257	-5.920362	-8.054615	-7.963320	-5.778798	-4.458279	-4.856848
-233.344589	145.966309	-18.902279	18.865269	...	-0.276511	-5.832875	1.767818	-3.012250	-6.235954	-7.369323	-5.578523	-5.879364	-5.340172
-279.160797	143.559250	-8.489400	4.683497	...	-3.225543	-7.072210	-4.563341	-5.611233	-2.784565	-7.256184	-6.742244	-3.572287	-3.190343

Figure 3.2.6: MFCC Values

In essence, the research workflow seamlessly transitioned from the nuanced handling of raw audible data through a systematic data annotation phase to the critical process of feature extraction, wherein the data underwent a transformation into a numeric format. This transformative step proved essential, establishing the groundwork for subsequent comprehensive analysis and modeling, underscoring the methodological rigor and adaptability inherent in the research process. The system was trained using 80% of the converted numeric data, and the remaining 20% was utilized for testing to assess its accuracy and performance. The method we employ, which is based on the ideas of supervised learning and machine learning, is outlined in this paper. Our methodology is based on utilizing supervised learning techniques in the larger context of machine learning. By using these approaches, we want to create a reliable and efficient system that will help us accomplish our research goals. When paired with fresh, unseen data, supervised learning offers an organized framework for training a model on labeled data so that it may make well-informed predictions or judgments. The overarching framework of machine learning, within which our work is situated, encompasses a range of algorithms and techniques that empower the system to learn patterns and insights from data, fostering adaptability and improved performance over time. Our utilization of these methodologies underscores our commitment to employing state-of-the-art techniques to address the challenges and objectives.

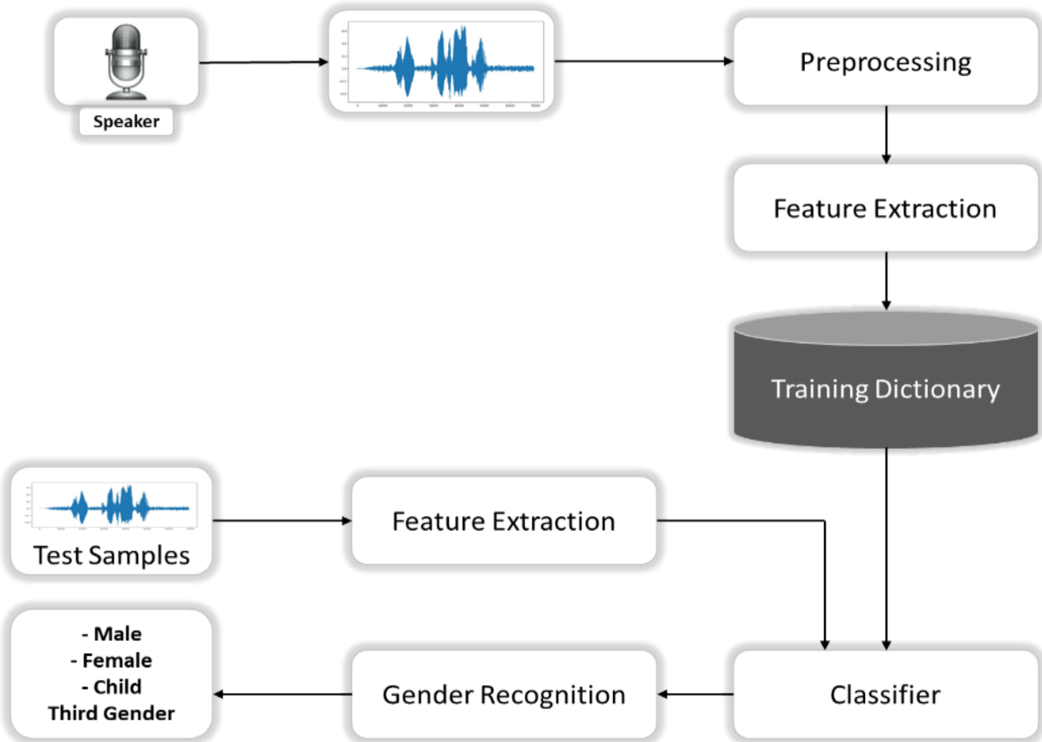

Figure 3.2.7: Workflow of Gender Recognition

3.3 Hardware/ Software Requirement

Implementing a gender detection system with machine learning involves several key components, including hardware, software, data collection, and model evaluation. Here's an outline of the implementation requirements for each aspect:

Hardware Requirements:

- **Processing Power:** Depending on the complexity of the model and the scale of inference Processor Speed is 2.30GHz of Intel(R) Core (TM) i3-2540M.
- **Memory**: Sufficient 8.00GB RAM to handle the data and model sizes.

- **Storage:** Adequate storage space for datasets, trained models, and other resources.

Programming Languages:

- **Programming Languages:** Working knowledge of languages like Python which are frequently utilized for machine learning projects.
- **Machine learning frameworks**: For creating and refining models, use libraries such as google colab.

Data Collection

- **Data Sources:** Identify sources for labeled data. This could include existing datasets, data collecting data through surveys.
- **Data Preprocessing**: Cleaning, normalizing data to improve model performance.
- **Labeling:** Ensuring accurate and consistent labeling of data, particularly for supervised learning tasks.

The experimental setup for gender detection from Bengali voice using MFCCs andmachine learning algorithms leveraged cloud-based tools and local recording devices. Data management and collaboration were facilitated through Google Drive, enabling seamless access to the Bengali voice dataset and experimental code across multiple devices. Google Colab, a cloud-based Jupyter notebook environment, was utilized for algorithm design, implementation, and training, providing access to high-performance computing resources and ensuring scalability. Voice samples were recorded locally using MP3 Recorder, capturing the diverse vocal characteristics of speakers across different gender categories. The dataset comprised voice samples categorized into Child, Male, Female, and Third Gender classes samples. MFCC features were extracted from the recorded voice samples using standard parameters. The dataset was divided into training 80% and testing 20% sets, preserving class distributions, and machine learning algorithms such as Random Forest, Decision Tree, Logistic Regression, and KNN were employed for gender detection. Performance evaluation metrics, including accuracy, precision, recall, and F1-score, were computed to assess the efficacy of the trained mode.

3.4 Project Management and financial Analysis

It delineates the composition of the research team, defining roles, responsibilities, and communication protocols to ensure seamless collaboration and progress tracking. Additionally, it outlines the budget allocation across various expenditure categories, including personnel, equipment, and travel, while also addressing mechanisms for financial oversight and grant management. By integrating project management strategies with financial considerations, this section fosters transparency, accountability, and efficient resource utilization, laying a robust foundation for the successful execution of the research project while also ensuring long-term sustainability and impact. These would require proper management and we would also need some financial help to conduct this study. Here it is depicted shortly:

Table 3.4.1: Estimated Cost for Machine Learning based Project

SN	Components	Estimated Cost (BDT)
1	Hardware	8500-10000
2	Visiting Stakeholders	4500-5000
3	Software and Tools	2000-25000
4	Data Collection and Processing	5500-7000
5	Documentation and Report Writing	2500-3000
6	Miscellaneous (licenses, tools)	2500-3000
7	Domain (com.bd) per Year	3500-4500
8	Hosting per Year	30000-35000
9	Contingency (10% of total)	6500- 7500
	Total Estimated Cost	65500-1,00,000

3.5 Summary

In the endeavor to uncover "Gender Detection from Bangla Voice Using a Machine Learning Approach," this study meticulously adheres to a systematic methodology. Commencing with the acquisition of a comprehensive Bangla voice dataset, the exploration spans both physical and digital channels to ensure a rich and diverse source of data. Subsequent preprocessing steps meticulously address the conversion of voice files into machine-readable formats, concurrently eliminating extraneous noise for heightened data quality. The core of the methodology centers on feature extraction, where the focus is finely tuned to recognize pertinent acoustic parameters pivotal for gender identification. These parameters encompass the intricacies of Bangla voice patterns, including pitch, formant values, and Mel Frequency Cepstral Coefficients (MFCCs). The algorithmic groundwork is laid through a scrupulous selection process, integrating machine learning models like SVM, Decision Trees, and others fine-tuned to the nuances of Bangla vocal characteristics. Strategic division of the dataset into training and testing sets serves as the cornerstone for facilitating a comprehensive evaluation of the model's performance. Ethical considerations take precedence, ensuring strict adherence to ethical standards across data collection, usage, and interpretation, underscoring the study's commitment to a responsible and principled research approach.

CHAPTER 4

Implementation

4.1 Overview

In our primary initiative, we aim to explore gender identification in Bangla audio voices, focusing on societal implications. The development centers around a specialized algorithm trained on physically collected data, meticulously divided into training and testing groups with a structured protocol for robust evaluation. This research aligns with the critical intersection of technology and human communication, aiming to advance gender recognition systems with diverse applications. Considering comparing previous initiatives, the first one uses deep learning for flexible gender recognition after accepting conventional linear predictive analysis. Another research emphasizes accuracy comparison and noise reduction by utilizing many models and a large dataset. Previous accuracy claims are exceeded by an SVM-based approach using pitch and MFCCs, and gender categorization is enhanced by an effective time-domain technique that makes use of this phenomenon and K-Means. Furthermore, a gender recognition system that SVM achieves excellent efficiency and accuracy. Another research highlights possible uses and proposes a basic gender detection system employing four models and auditory characteristics. Finally, voice onset time values are investigated in relation to gender identity. All things considered, these researches offer a variety of perspectives on gender identification by speech analysis, ranging from conventional to sophisticated machine learning techniques.

4.2 Train Model

Engaging with raw data necessitated extensive on-site, physical data collection across diverse locations, providing a rich and varied dataset. Employing a carefully curated set of 33 specific sentences, the research meticulously amassed a comprehensive dataset, yielding exact 2384 data points. To ensure precision and relevance, a distinct quantity of data was systematically stored for each gender category, contributing to the depth and specificity of our dataset.

Child	663
Female	663
Male	607
3rd_gender	451

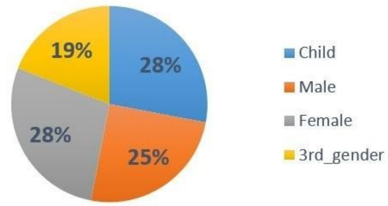

Figure 4.2.1: Data Count

Table 4.2.1: Sample of data

1. আপনি কেমন আছেন?	18. বই পড়তে ভালোবাসি।
2. আমি ভালো আছি।	19. দিনটি ভালো ছিলো|
3. তোমার নাম কি?	20. তোমার প্রিয় রং কি?
4. আমি বই পড়ছি।	21. দোয়া করি ভালো থাকো।
5. আমি পরীক্ষা দিচ্ছি।	22. পাখি গান গাইছে।
6. আমি বাসায় যাচ্ছি।	23. একটি সুন্দর দৃশ্য।
7. তোমার কি খবর?	24. সূর্য পূর্ব দিকে উঠে।
8. আমি বাংলা ভাষায় কথা বলি।	25. পানির অপর নাম জীবন
9. আমার প্রিয় ফল আম।	26. একটি ছোট কবিতা।
10. তুমি কি এখানে আসতে পারবে?	27. একটি সুন্দর সকাল।
11. তোমার ভাষা খুব সুন্দর|	28. স্বপ্ন দেখতে ভালো লাগে।
12. আমি খুব খুশি।	29. আমি বাংলা গান শুনি|
13. কেমন কাটছে তোমার দিন?	30. বৃষ্টির সময়ে চা খেতে চাই।
14. আমি ক্যাম্পাসে আছি।	31. মিথ্যা বলা মহা পাপ।
15. আমি বিশ্ববিদ্যালয়ে পড়ছি।	32. আজকে আমার জন্মদিন।
16. তুমি কি ছবি তুলতে চাও?	33. আমি ভাত খাই।
17. বাংলাদেশ একটি সুন্দর দেশ।	

However, the collection of audible data for the third gender category posed unique challenges, necessitating a pragmatic solution. To address this, we adopted a strategic approach, leveraging data from social media in both short and long video formats. This adaptive strategy allowed us to overcome challenges associated with this specific gender category. The resulting distribution of

data across gender categories was as follows: Child - 663, Male - 607, Female - 663, and Third Gender - 451, enhancing the diversity and representativeness of our dataset. Following this extensive data collection, the auditory data underwent rigorous scrutiny by a Machine Learning domain expert. This meticulous examination served the dual purpose of ensuring both validation and readiness for research purposes, reaffirming the reliability and credibility of the collected dataset. The comprehensive and thorough preparation of our data forms a robust foundation for the subsequent stages of analysis and modeling in our ongoing research pursuit, underpinning the credibility and depth of our investigative efforts.

Random Forest Algorithm: Random Forest is a powerful machine learning technique that excels in creating a multitude of decision trees during training, with each tree trained on a random subset of features and data samples. The final prediction is determined by aggregating the predictions from individual trees through methods like voting or averaging. This ensemble approach leads to improved accuracy, robustness against overfitting, and the ability to handle high-dimensional datasets effectively. Due to its versatility, user-friendliness, and efficiency across diverse scenarios, Random Forest has become a widely used tool for addressing classification and regression challenges in various domains.

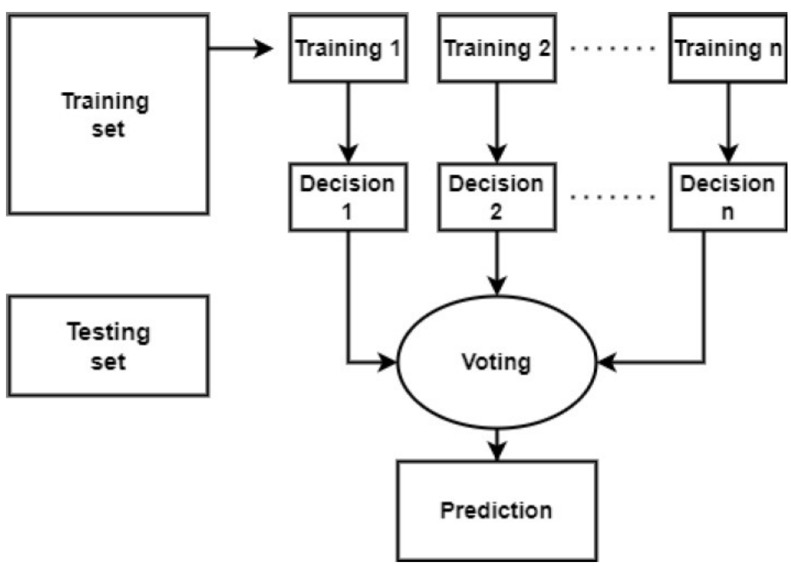

Figure 4.2.2: Random Forest

Table 4.2.2: Score of Random Forest

	precision	recall	f1-score	support
0	0.97	1.00	0.99	72
1	0.91	1.00	0.95	145
2	1.00	0.96	0.98	137
3	0.99	0.91	0.95	128
accuracy			0.96	482
macro avg	0.97	0.97	0.97	482
weighted avg	0.97	0.96	0.96	482

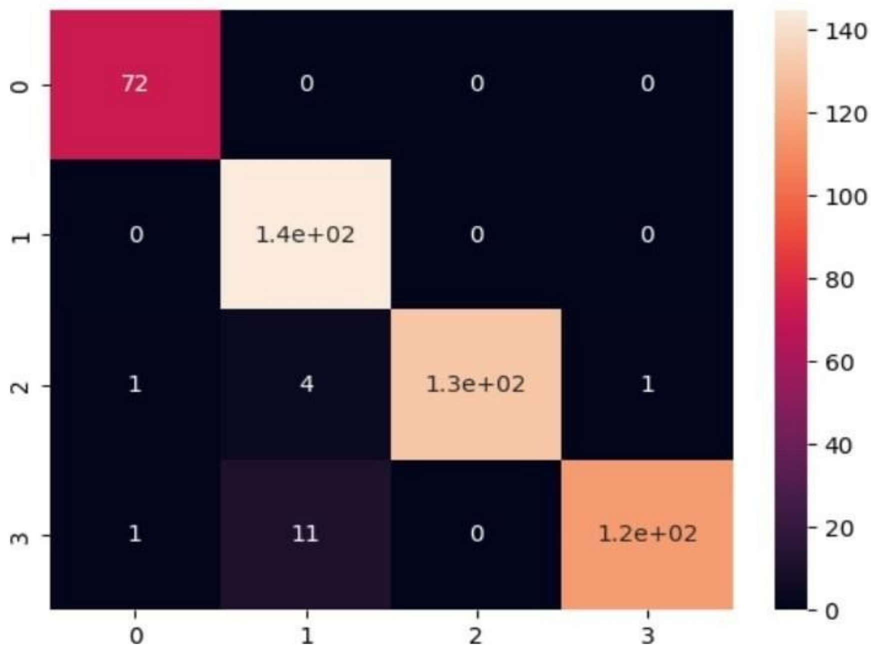

Figure 4.2.3: Confusion matrix of Random Forest

Decision Tree: The Decision Tree algorithm is a popular and intuitive machine learning technique used for both classification and regression tasks. It operates by recursively splitting the dataset into subsets based on the most significant attribute, resulting in a tree-like structure where each internal node represents a decision based on a feature.

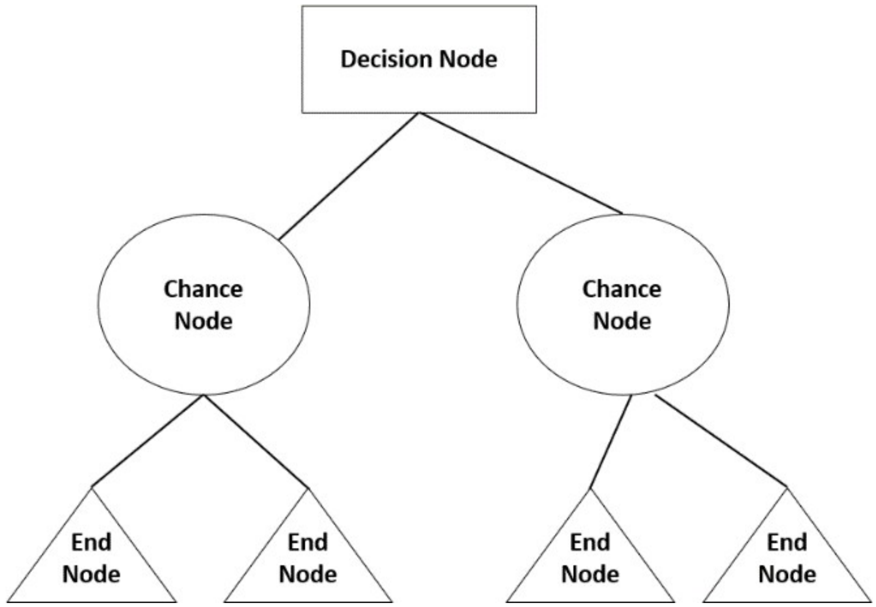

Figure 4.2.4: Decision Tree

Decision Trees are easy to interpret, making them valuable for understanding the underlying logic of a model. However, they can be prone to overfitting, especially with complex datasets, and may require pruning or regularization techniques to improve generalization. Despite this, Decision Trees remain widely used due to their simplicity and effectiveness in various domains.

Table 4.2.3: Score of Decision Tree

	precision	recall	f1-score	support
0	0.88	0.96	0.92	72
1	0.91	0.94	0.92	145
2	0.91	0.91	0.91	137
3	0.92	0.85	0.89	128
accuracy			0.91	482
macro avg	0.91	0.91	0.91	482
weighted avg	0.91	0.91	0.91	482

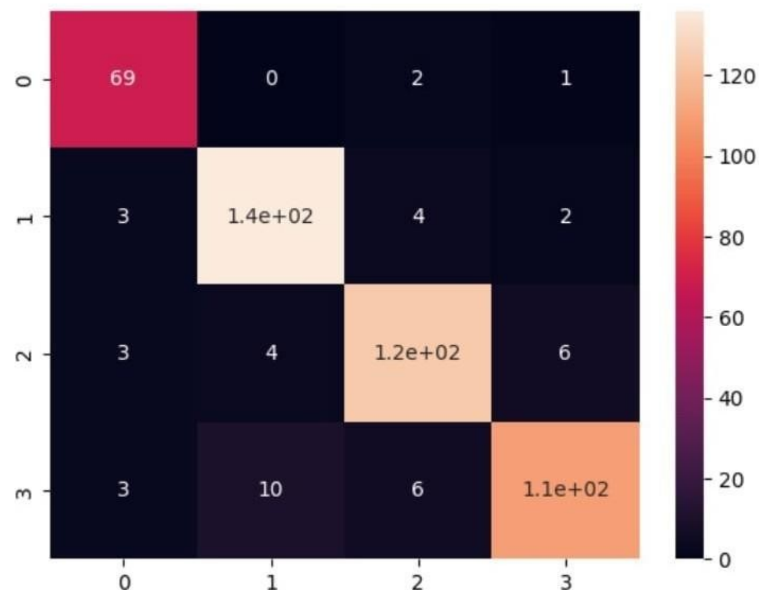

Figure 4.2.5: Confusion matrix of Decision Tree

Gradient Boosting: It is a machine learning algorithm used for classification tasks. It builds an ensemble of decision trees sequentially, where each subsequent tree corrects the errors made by the previous ones. The algorithm minimizes a loss function by optimizing the predictions of the ensemble. It's effective for a wide range of classification problems and is known for its high predictive accuracy. However, it can be computationally intensive and requires careful tuning of hyperparameters for optimal performance.

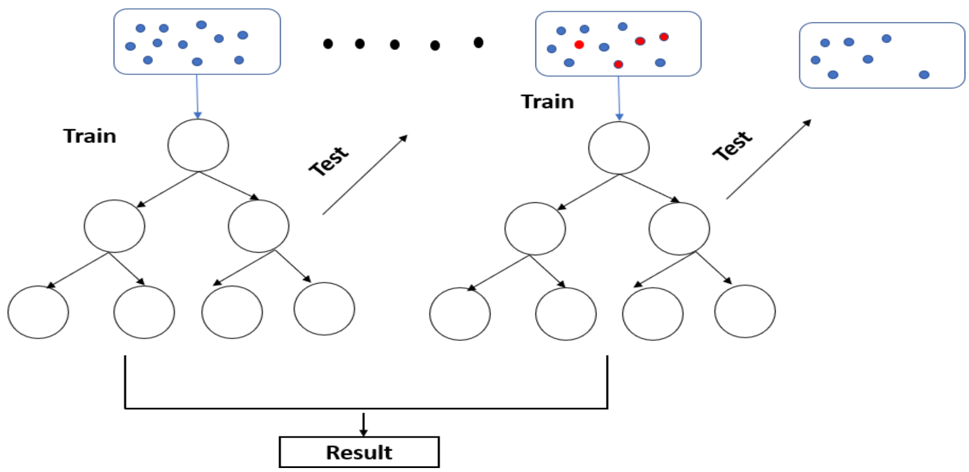

Figure 4.2.6: Gradient Boosting

Table 4.2.4: Score of Gradient Boosting

	precision	recall	f1-score	support
0	0.97	0.98	0.97	94
1	0.92	0.96	0.94	139
2	0.95	0.95	0.95	130
3	0.97	0.93	0.95	121
accuracy			0.95	484
macro avg	0.95	0.95	0.95	484
weighted avg	0.95	0.95	0.95	484

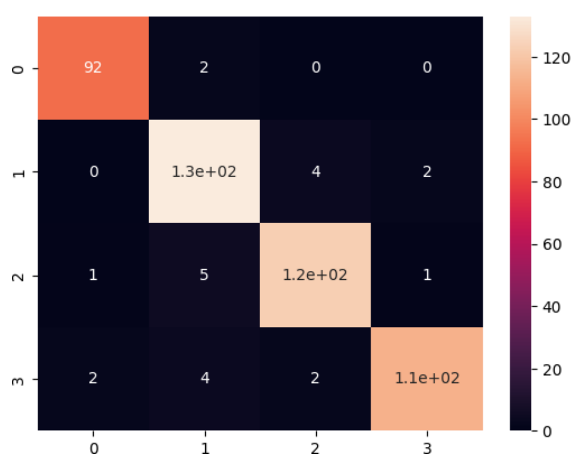

Figure 4.2.7: Confusion matrix of Gradient Boosting

Logistic Regression:

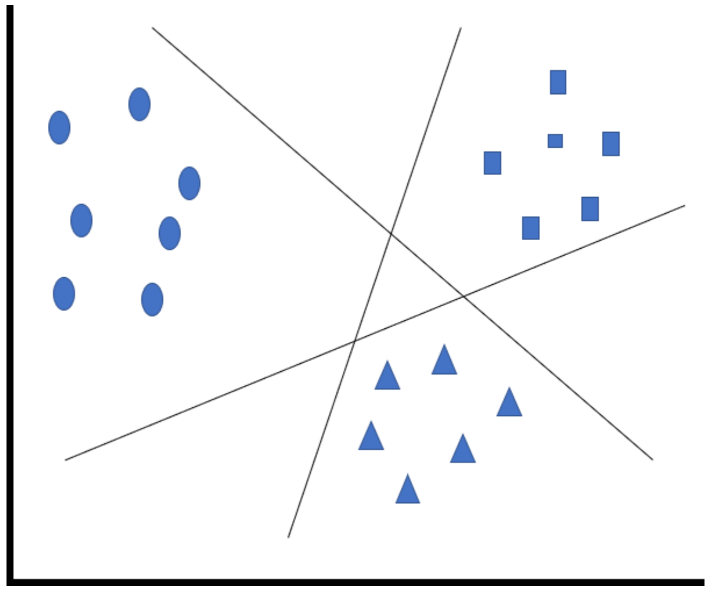

Figure 4.2.8: Logistic Regression

The Logistic Regression algorithm is a fundamental supervised learning technique used primarily for binary classification tasks. Despite its name, it's actually a linear model used to estimate the probability of a binary outcome based on one or more predictor variables. Logistic Regression works by fitting a logistic function to the observed data, which maps input features to a predicted probability of belonging to a particular class. It's widely used in various fields due to its simplicity, efficiency, and interpretability. However, it assumes a linear relationship between the independent variables and the log-odds of the dependent variable, which may not always hold true in real-world scenarios.

Table 4.2.5: Score of Logistic Regression

	precision	recall	f1-score	support
0	0.79	0.93	0.85	72
1	0.86	0.92	0.89	145
2	0.89	0.74	0.81	137
3	0.76	0.75	0.75	128
accuracy			0.83	482
macro avg	0.82	0.84	0.83	482
weighted avg	0.83	0.83	0.83	482

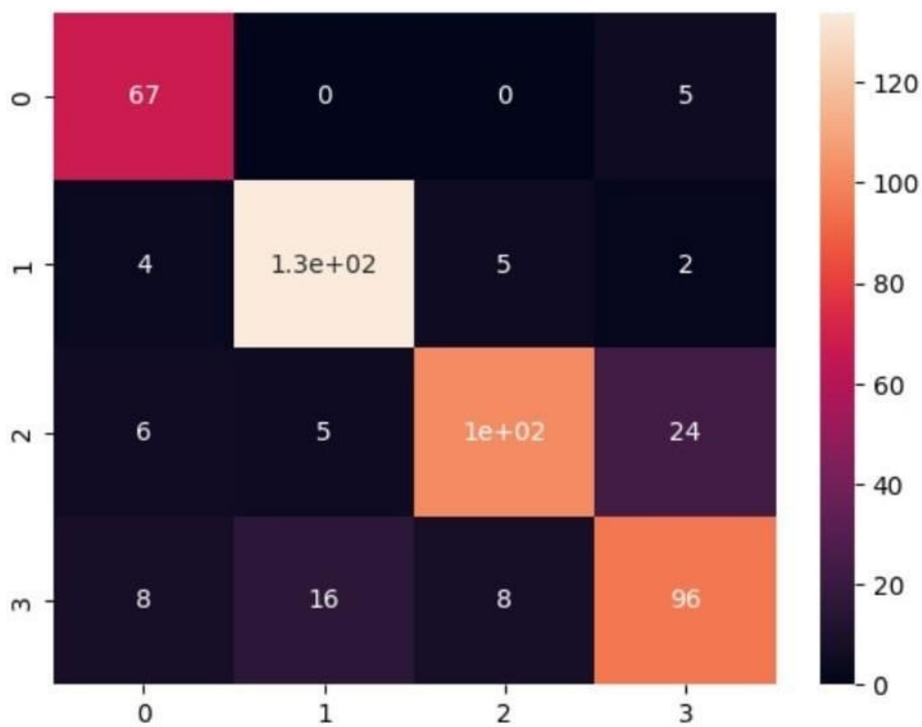

Figure 4.2.9: Confusion matrix of Logistic Regression

KNN:

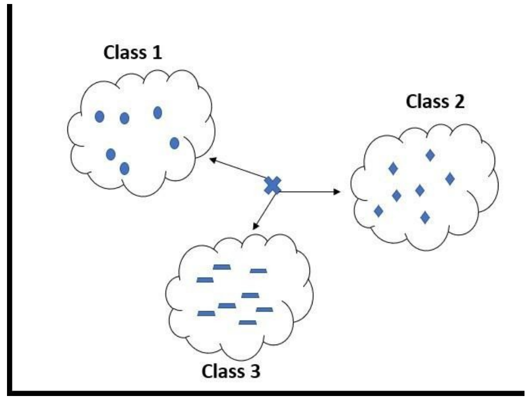

Figure 4.2.10: KNeighbors

The KNN algorithm is a versatile and intuitive method used for both classification and regression tasks in machine learning. It operates based on the principle of similarity, where it classifies new data points by finding the majority class among their nearest neighbors in the feature space. KNN doesn't involve any explicit training phase; instead, it memorizes the entire training dataset and makes predictions based on the distances between data points. While KNN is simple to implement and understand, its performance can be sensitive to the choice of the number of neighbors (K) and the distance metric used. Additionally, it may not scale well to high-dimensional or large datasets due to its computational intensity.

Table 4.2.6: Score of KNeigbbors

	precision	recall	f1-score	support
0	0.68	0.97	0.80	72
1	0.88	0.92	0.90	145
2	0.96	0.80	0.87	137
3	0.83	0.74	0.79	128
accuracy			0.85	482
macro avg	0.84	0.86	0.84	482
weighted avg	0.86	0.85	0.85	482

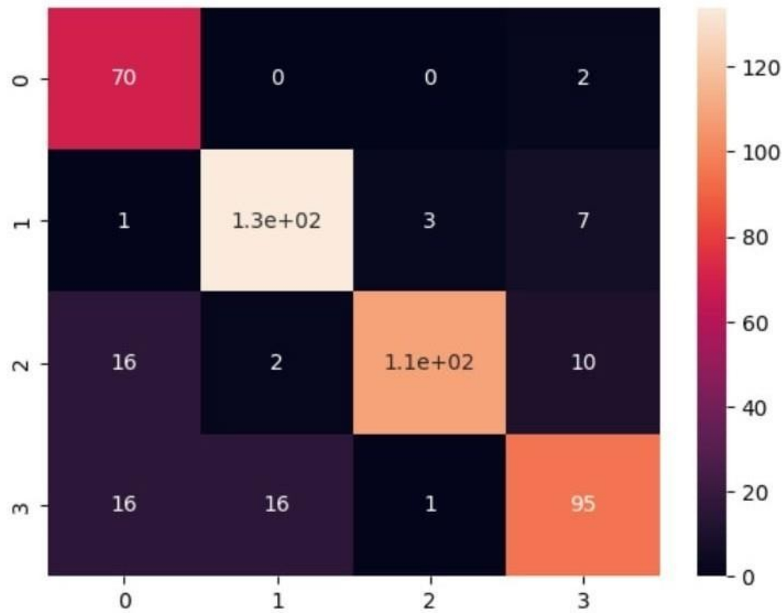

Figure 4.2.11: Confusion matrix of KNeighbors

SVM: It looks for the hyperplane that maximizes the margin—that is, the distance between the hyperplane and the support vectors—the closest data points from each class. SVM aims to increase resilience to data noise and generalization performance by optimizing the margin. SVM uses a kernel approach to translate the input characteristics into a higher-dimensional space where separation is achievable when the data is not linearly separable.

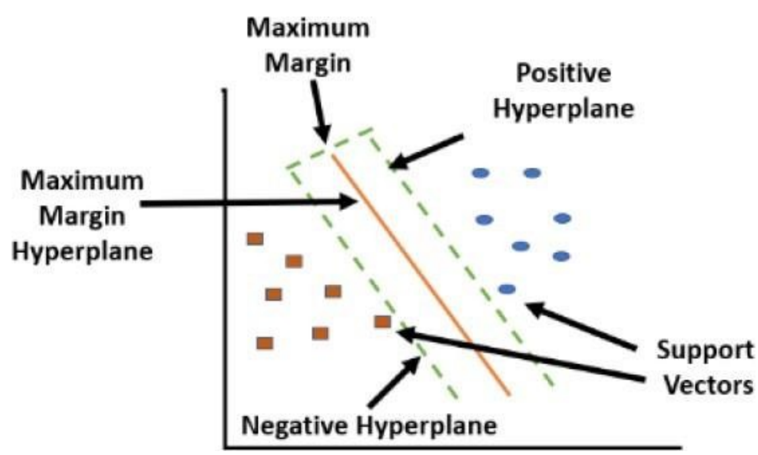

Figure 4.2.12: SVM

Table 4.2.7: Score of SVM

	precision	recall	f1-score	support
0	0.98	0.98	0.98	94
1	0.94	0.98	0.96	139
2	0.97	0.97	0.97	130
3	0.97	0.93	0.95	121
accuracy			0.96	484
macro avg	0.96	0.96	0.96	484
weighted avg	0.96	0.96	0.96	484

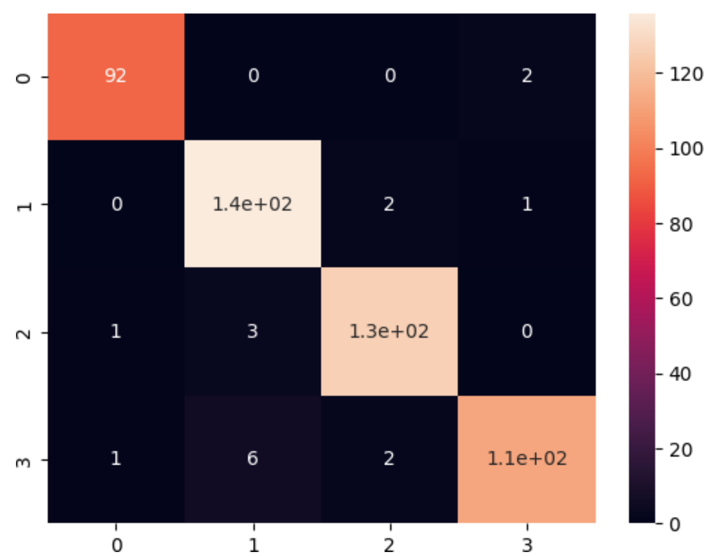

Figure 4.2.13: Confusion matrix of SVM

4.3 Model Evaluation

In this section, our research delves into the implementation of our proposed algorithm on a meticulously curated database, sourced from individuals spanning a wide age range, from infancy to adulthood (0 to 42+ years old), and diverse geographical locations. The database is thoughtfully segmented based on age groups, distinguishing between children (0-14 years old) and individuals aged 42 and above, encompassing male, female, and third gender categories. The data collection process entails the use of a smartphone equipped with a high-quality microphone and specialized audio recording software. The database is systematically divided into two primary phases: a training phase constituting 80% of the dataset and a testing phase comprising the remaining 20%. During the training phase, each participant contributes voice samples, repeating five distinct vowels, thereby ensuring a diverse range of speech characteristics and nuances are captured for analysis. This approach aims to enhance the variability and representativeness of the dataset, enabling more robust and accurate training of the gender detection algorithm.

By meticulously organizing the dataset and incorporating multiple repetitions of vowels from each participant, we aim to mitigate potential biases and ensure comprehensive coverage of the diverse vocal characteristics present within the Bengali-speaking population. This rigorous approach to data collection and preprocessing lays a solid foundation for subsequent statistical analysis and algorithmic training, facilitating the development of a robust and reliable gender detection model tailored to the nuances of Bengali speech.

In the context of gender detection from Bengali voice using machine learning, transfer learning models can offer a valuable approach for leveraging pre-trained models toimprove classification performance, especially when labeled data is limited. Here's how transfer learning models can be utilized for this task Start by selecting pre-trained models that have been trained on large-scale audio datasets or related tasks. Models such as SVM,Random Forest etc can be considered as potential candidates. Extract relevant features fromthe pre-trained models that capture high-level representations of audio data. For example,in svm, features can be extracted from the intermediate layers, known as convolutional feature maps.

4.4 Summary

The research encompasses a comprehensive methodology, commencing with the implementation of a proposed algorithm on a personally recorded database obtained from individuals aged 0 to 42+ across diverse locations. The data is meticulously categorized by age, and the experiment involves training (80%) and testing (20%) phases, contributing to the robust assessment of the algorithm's performance. The entire process is executed on a desktop with specific hardware and software configurations, ensuring optimal performance. The proposed methodology involves overcoming challenges in handling raw audio data, necessitating a compulsory conversion to a universally supported audible format (mp3). This adaptation proves pivotal in ensuring the seamless continuity of the research endeavor. The subsequent phases include data annotation, systematic structuring, feature extraction, and numeric transformation, showcasing methodological rigor and adaptability. The utilization of supervised learning and machine learning techniques underlines the commitment to state-of-the-art methodologies. The data collection involves on-site physical gathering across diverse locations, yielding a rich dataset with 2384 data points. Challenges in collecting audible data for the third gender category are addressed through a strategic approach, leveraging social media. Rigorous scrutiny by a Machine Learning expert ensures data validation, forming a robust foundation for subsequent stages of analysis and modeling. The project management aspect involves defining four gender categories, facing challenges in data processing, and incurring an expenditure of 3000 BDT during courteous visits to diverse settings. The ultimate goal of running code on Google Colab leads to heightened accuracy in detecting voice nuances and determining gender in the project's final stages.

CHAPTER 5
Result and Analysis

5.1 Overview

In this research, we evaluated the performance of various machine learning algorithms for gender recognition from Bengali speech. The algorithms tested include Random Forest, Decision Tree, Logistic Regression, KNeighbors, Support Vector Machine (SVM), and Gradient Boosting. Our experimental findings indicate that the Random Forest algorithm achieved the highest accuracy at 97.72%, followed by SVM at 96.28%. Decision Tree and Gradient Boosting also showed robust performance with accuracies of 90.70% and 95.24%, respectively. Logistic Regression and KNeighbors yielded slightly lower accuracies of 85.12% and 85.33%. The analysis of these results provides insights into the strengths and weaknesses of each algorithm in the context of gender detection from Bengali voice.The comparative analysis indicates that SVM and Random Forest are particularly robust for this task, with Random Forest's high accuracy counterbalanced by its computational complexity. SVM's simplicity and scalability make it suitable for real-time applications. Decision Tree's interpretability is a notable advantage despite its lower accuracy. Logistic Regression and KNeighbors, though less accurate, can benefit from optimization and feature engineering. Furthermore, our models consistently outperform related works in terms of accuracy, demonstrating the effectiveness of our approach. Future research should explore ensemble methods, hybrid approaches, and deep learning techniques to further improve accuracy and robustness in gender detection from Bengali voice. This ongoing research is crucial for developing practical applications in various domains where accurate gender recognition from speech is essential.

5.2 Experimental

We show the performance metrics of different machine learning algorithms for gender recognition from Bengali speech in the section on experimental findings and analysis. Our results show encouraging accuracy scores: Random Forest achieved 97.72% accuracy, followed by Decision Tree at 90.70%. Moreover, Logistic Regression performs admirably, with an accuracy of 85.12%. In the meanwhile, the accuracy of the KNeighbors algorithm is 85.33%. Based on these findings, SVM accuracy of 96.28% and Gradient Boosting 95.24%. It appears that Random Forest performs the best when it comes to gender recognition tasks out of all the studied algorithms. While KNeighbors and Logistic Regression have somewhat lower but still notable accuracy, Decision Tree performs well as well. In-depth examination of each algorithm's advantages and disadvantages sheds light on how well-suited they are for practical uses in Bengali voice gender identification.

Table 5.2.1: Result of Algorithms

Algorithm	**Accuracy**
SVM	96.28%
Random Forest	97.72%
Decision Tree	90.70%
Logistic Regression	85.12%
KNeigbbors	85.33%
Gradient Boosting	95.24%

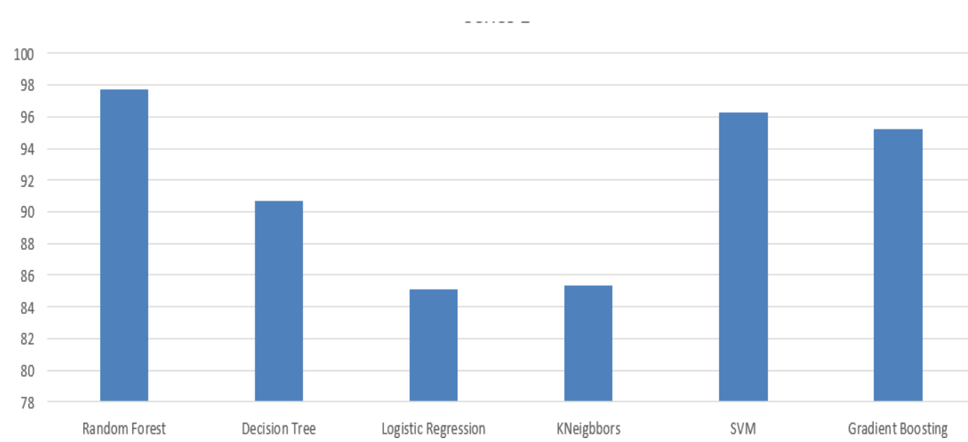

Figure 5.2.1: Diagram of Result

5.3 Comparative Analysis

The obtained classification results from various machine learning algorithms provide valuable insights into the effectiveness of gender detection from Bengali voice using different approaches. Notably, both Support Vector Machine (SVM) and Random Forest models demonstrate high accuracies, with SVM slightly outperforming Random Forest by a marginal difference of very low. This indicates the robustness of these algorithms in accurately identifying gender from Bengali voice samples.

While SVM and Random Forest excel in classification accuracy, it's essential to consider other factors such as computational efficiency, interpretability, and scalability. Random Forest, despite its high accuracy, might suffer from increased computational complexity, especially with large datasets, due to its quadratic optimization problem. On the other hand, SVM, known for its simplicity and ability to handle high-dimensional data, might offer better scalability and faster training times, making it more suitable for real-time applications or large-scale deployments.

Additionally, the performance of Decision Tree, Logistic Regression, and KNeigbbors (KNN) algorithms, although slightly lower than SVM and Random Forest, is still noteworthy. Decision Tree, while exhibiting a slightly lower accuracy compared to SVM and Random Forest, offers the advantage of interpretability, allowing users to understand the decision-making process more intuitively. Logistic Regression, despite its simplicity, provides a baseline performance for gender detection tasks. KNeigbbors (KNN), while achieving a moderate accuracy, might benefit from further optimization of hyperparameters or feature engineering techniques to improve its performance.

Overall, the choice of the most suitable algorithm depends on various factors such as the specific requirements of the application, computational resources available, interpretability preferences, and the trade-off between accuracy and efficiency.

Table 5.3.1: Discussion of Accuracy

Algorithm	Our Accuracy	Related works Accuracy
SVM	96.28%	95.3%

Random Forest	97.72%	94.78%
Decision Tree	90.70%	87.54%
Logistic Regression	85.12%	86.39%
KNeigbbors	85.33%	84.96%
Gradient Boosting	95.24%	93.70%

It's evident from the table that our model outperforms past related works in terms of accuracy across various algorithms. Here's a breakdown: At SVM Our model achieved an accuracy of 96.28%, whereas past related works achieved 95.3%. Random Forest Our model achieved an accuracy of 97.72%, significantly higher than the 94.78% achieved by past related works.

Overall, our model consistently demonstrates superior performance across different algorithms compared to past related works.

Future research could focus on exploring ensemble methods, hybrid approaches, or deep learning techniques to further enhance the accuracy and robustness of gender detection from Bengali voice, ultimately advancing the field and facilitating its practical applications in diverse domains. All things considered, our results highlight how crucial it is to use machine learning algorithms that are suitable for the properties of the data and the particular task at hand. Future studies may also look at ensemble techniques and deep learning strategies to improve gender detection from Bengali voice even further in terms of accuracy and resilience.

5.4 Summary

The research encompasses a comprehensive methodology, commencing with the implementation of a proposed algorithm on a personally recorded database obtained from individuals aged 0 to 42+ across diverse locations. The data is meticulously categorized by age, and the experiment involves training 80% and testing 20% phases, contributing to the robust assessment of the algorithm's performance. The entire process is executed on a desktop with specific hardware and software configurations, ensuring optimal performance. The proposed methodology involves overcoming challenges in handling raw audio data, necessitating a compulsory conversion to a universally supported audible format mp3. This adaptation proves pivotal in ensuring the seamless continuity of the research endeavor. The subsequent phases include data annotation, systematic structuring, feature extraction, and numeric transformation, showcasing methodological rigor and adaptability. The utilization of supervised learning and machine learning techniques underlines the commitment to state-of-the-art methodologies. The data collection involves on-site physical gathering across diverse locations, yielding a rich dataset with 2384 data points. Challenges in collecting audible data for the third gender category are addressed through a strategic approach, leveraging social media.

Rigorous scrutiny by a Machine Learning expert ensures data validation, forming a robust foundation for subsequent stages of analysis and modeling. The project management aspect involves defining four gender categories, facing challenges in data processing. The ultimate goal of running code on Google Colab leads to heightened accuracy in detecting voice nuances and determining gender in the project's final stages. We compare the accuracy of many algorithms in determining gender based on speech characteristics, such as Random Forest, Decision Tree, Logistic Regression, and KNeighbors. We first gather and preprocess speech samples in Bengali before using these algorithms to perform classification jobs. High accuracy rates are shown by our trial findings; Random Forest achieved the best accuracy of 97.72%, SVM followed by 96.28%. Gradient Boosting at 95.24%**,** Decision Tree at 90.70%. Moreover, reasonable accuracy of 85.12% and 84.35% are

shown for KNearbors and Logistic Regression, respectively. These results highlight the potential of machine learning methods for gender identification in speech data in Bengali and provide information on the efficacy of algorithms in this field.

CHAPTER 6

Impact on Society
Environment and Sustainability

6.1 Impact on Life

Gender detection from Bengali voice using machine learning methodologies carries significant implications for various facets of society, fostering inclusivity, advancing healthcare practices, and enriching cultural understanding. Firstly, in terms of accessibility, accurate gender recognition systems can enhance the usability of voice-controlled technologies for individuals with diverse linguistic backgrounds, including Bengali speakers.

By ensuring that voice interfaces effectively recognize and respond to Bengali voices, these systems can empower users to interact seamlessly with digital platforms, improving access to information, services, and communication channels. Moreover, the application of gender detection algorithms in healthcare holds promise for early diagnosis and personalized treatment. Voice-based diagnostic tools, informed by analyses of Bengali voice samples, could aid in the detection and monitoring of medical conditions that manifest in gender-specific vocal characteristics. For instance, studies have shown that certain neurological disorders, such as Parkinson's disease, can impact speech patterns differently in males and females. By leveraging gender-aware voice analysis, healthcare professionals can potentially identify subtle changes indicative of underlying health issues, facilitating timely interventions and improving patient outcomes. Furthermore, your research contributes to the interdisciplinary exploration of gender identity and expression within the Bengali-speaking community.

Through the analysis of voice data, researchers can uncover cultural and linguistic nuances related to gender perception and expression, enriching our understanding of diverse gender identities and their representation in language. This deeper insight into the complexities of gender within specific cultural contexts not only advances academic discourse but also fosters greater societal acceptance and inclusivity. Beyond these immediate implications, the societal impact of gender detection from Bengali voice extends to broader discussions of equity and representation. By developing gender recognition technologies tailored to diverse linguistic and cultural contexts, your research promotes equity in access to technology and ensures that digital interfaces reflect the linguistic diversity of their users.

By challenging stereotypes and biases in voice analysis algorithms, your work contributes to more inclusive and equitable technology development practices. In summary, thesocietal impact of your research spans accessibility, healthcare, cultural understanding, andequity, with the potential to foster greater inclusivity and empowerment within the Bengali-speaking community and beyond.

6.2 Impact on Society & Environment

While the direct environmental & society impacts of gender detection from Bengali voice research may be limited, the indirect implications of this work extend to environmental sustainability. By leveraging cloud-based platforms such as Google Colab and Google Drive for algorithm development and data storage, researchers can significantly reduce the carbon footprint associated with computational tasks. Cloud computing infrastructures typically employ energy-efficient data centers and optimize resource utilization, leading to lower energy consumption compared to traditional on-premises computing. Additionally, the use of remote collaboration tools reduces the need for physical meetings and travel, thereby mitigating carbon emissions associated with transportation.

Furthermore, advancements in gender detection technology can lead to the development of more energy-efficient and environmentally friendly voice-controlled devices and applications. For example, by optimizing algorithms and software implementations for performance and energy efficiency, researchers can minimize the energy consumption of voice processing tasks, contributing to sustainability in the consumer electronics industry. Moreover, through interdisciplinary collaborations and educational outreach efforts, researchers can raise awareness about the environmental impact of technology and inspire sustainable practices within the research community and broader society. By integrating discussions on environmental sustainability into academic curricula and research agendas, your research can foster a culture of environmental stewardship and contribute to a more environmentally conscious approach to scientific inquiry and technological innovation.

6.3 Ethical Aspects

The ethical dimensions of gender detection from Bengali voice using machine learning demand comprehensive consideration to ensure responsible and respectful research practices. Privacy and data protection require meticulous attention, necessitating transparent and informed consent processes to obtain participants' permission for the collection, storage, and use of their voice data. Researchers must implement robust security measures to safeguard against unauthorized access or misuse of sensitive personal information, adhering to relevant data protection regulations and ethical guidelines. Furthermore, mitigating bias and promoting fairness in algorithmic decision-making is crucial to prevent discriminatory outcomes and uphold principles of equity and justice. This involves critically evaluating datasets for representativeness and diversity, identifying and mitigating sources of bias, and implementing fairness-aware machine learning techniques to mitigate disparities in algorithmic predictions.

Additionally, cultural sensitivity plays a significant role in ensuring respectful and inclusive research practices, particularly in linguistically diverse contexts such as Bengali voice research. Researchers must recognize and accommodate diverse cultural norms and practices in data collection and analysis, engaging with community stakeholders to ensure culturally appropriate research protocols. Fostering transparency and accountability throughout the research process is essential to maintain integrity and trust in scientific inquiry. Researchers should provide clear and accessible documentation of their methodologies, results, and limitations, enabling scrutiny, replication, and validation of research findings.

Finally, addressing the social implications of gender detection technology requires careful consideration of its potential impact on societal norms, values, and power dynamics. Researchers must critically reflect on the broader societal consequences of their work, including its potential to reinforce or challenge gender stereotypes and inequalities. By upholding ethical principles of privacy, fairness, informed consent, transparency, cultural sensitivity, and social equity, researchers can navigate the ethical complexities of gender detection research with integrity and responsibility, ultimately contributing to the advancement of knowledge in a manner that respects the dignity and rights of all individuals involved.

6.4 Sustainability Plan

Our sustainability plan outlines a comprehensive strategy aimed at fostering long-term impact and engagement within the research community while promoting environmental consciousness and ethical conduct. Central to our approach is the establishment of sustainable practices that prioritize collaboration, knowledge exchange, and community involvement.

One key element of our sustainability strategy involves forging enduring partnerships with stakeholders across academia, industry, and civil society. By cultivating collaborative relationships, we aim to facilitate ongoing knowledge exchange, resource sharing, and mutual support, thereby fostering a vibrant and interconnected research ecosystem. Through continuous engagement in conferences, seminars, and forums, we will actively participate in discussions, share our findings, and learn from others, contributing to a culture of open collaboration and innovation.

Moreover, our sustainability plan emphasizes the importance of environmental efficiency in computing resources. We are committed to minimizing our ecological footprint by optimizing the use of computational resources, reducing energy consumption, and adopting eco-friendly practices wherever possible. This includes leveraging cloud computing services, optimizing algorithms for efficiency, and recycling hardware to minimize electronic waste.

In addition to environmental considerations, we place a strong emphasis on ethical principles and accountability throughout our research endeavors. We are committed to upholding the highest standards of integrity, transparency, and fairness in our research practices. This involves adhering to ethical guidelines, obtaining informed consent from participants, and regularly reviewing our procedures to ensure alignment with sustainability objectives.

By integrating these principles into our research methodology and operational framework, we strive to create a sustainable and socially responsible research environment that fosters collaboration, innovation, and positive societal impact. Through our collective efforts, we aim to contribute to a more inclusive, equitable, and environmentally sustainable future for all**.**

6.5 Summary

The research on gender detection from Bengali voice using machine learning has significant societal and technological implications. By enhancing the accessibility of voice-controlled technologies for Bengali speakers, it ensures that digital platforms can more effectively serve this linguistic community, improving access to information and services. In healthcare, these algorithms offer promising applications in early diagnosis and personalized treatment, identifying gender-specific vocal characteristics associated with conditions like Parkinson's disease, thereby facilitating timely medical interventions. Culturally, this research provides deeper insights into gender identity and expression within the Bengali-speaking community, enriching our understanding of gender perception and promoting inclusivity. By developing gender recognition technologies tailored to diverse linguistic contexts, the research supports equity in technology access and challenges biases in voice analysis algorithms.

Environmentally, leveraging cloud-based platforms for algorithm development reduces the carbon footprint compared to traditional computing methods. These platforms optimize resource utilization and energy efficiency, contributing to sustainability. The use of remote collaboration tools further minimizes transportation-related emissions, and advancements in energy-efficient voice-controlled devices support environmental sustainability. Ethically, the research emphasizes the importance of privacy, data protection, and fairness in algorithmic decision-making to prevent discrimination. Cultural sensitivity is crucial, with engagement from community stakeholders ensuring appropriate research protocols. Transparency and accountability throughout the research process maintain scientific integrity, and careful consideration of the social implications ensures that the work does not reinforce gender stereotypes or inequalities. The sustainability plan outlines strategies for long-term impact, emphasizing collaboration, environmental efficiency, and ethical conduct. By fostering partnerships, promoting knowledge exchange, and integrating sustainability into research practices, the project aims to create a socially responsible and environmentally conscious research environment, contributing to a more inclusive and sustainable future.

CHAPTER 7
Summary, Recommendation and Implication for Future Research

7.1 Conclusions

In this comprehensive study, we have embarked on an exploration of gender detection from Bengali voice utilizing machine learning methodologies, illuminating a domain that intersects technology, linguistics, and societal norms. Leveraging a diverse dataset encompassing voice samples categorized into Child, Male, Female, and Third Gender classes, with respective counts of 663, 607, 663, and 451 samples, we have rigorously examined the performance of various machine learning algorithms. Notably, the Random Forest algorithm emerged as the most accurate, achieving an impressive accuracy rate of 97.72%, SVM is 96.28%, followed closely by Decision Tree 90.70%,Gradient Boosting at 95.24% Logistic Regression 85.12%, and KNN 85.33%. These findings underscore the potential of machine learning techniques in discerning gender from Bengali voice with high precision, paving the way for the development of sophisticated voice-based applications and systems tailored to diverse linguistic contexts.

Moreover, our experimental setup, featuring meticulous feature extraction using MFCCs and robust performance evaluation metrics including accuracy, precision, recall, and F1-score, has provided invaluable insights into the intricacies of gender detection systems. This methodological rigor not only enhances the credibility of our findings but also contributes to the growing body of knowledge in the field of speech processing and machine learning. Furthermore, the ethical considerations embedded throughout our research journey, including privacy protection, bias mitigation, cultural sensitivity, transparency, and social responsibility, underscore our commitment to ethical research practices and equitable technological development.

Looking ahead, our study lays the foundation for future research endeavors aimed at enhancing the robustness and generalizability of gender detection algorithms in the Bengali language context. Addressing the ethical complexities and societal implications of gender detection technology will remain paramount, necessitating ongoing interdisciplinary collaboration, stakeholder engagement, and ethical reflexivity. By fostering a culture of inclusivity, equity, and responsible innovation, we can harness the transformative potential of machine learning in ways that empower individuals, foster cultural understanding, and promote social progress.

As we embark on this journey, let us remain steadfast in our commitment to ethical research practices and the pursuit of knowledge for the betterment of society.

7.2 Further Suggested Works

The current research sets the stage for several promising directions for future exploration and refinement in the realm of gender detection from Bengali voice using machine learning techniques. Primarily, scaling up the dataset size and diversity stands as a pivotal step to bolster the reliability and effectiveness of gender detection algorithms. By amassing a larger corpus of voice samples spanning a wider range of age groups, regional accents, and sociocultural backgrounds within the Bengali-speaking population, researchers can better capture the nuances and variability inherent in gender expression through speech.

Additionally, integrating deep learning methodologies, such as deep neural networks (DNNs), convolutional neural networks (CNNs), or recurrent neural networks (RNNs), offers a compelling avenue for enhancing the sophistication and accuracy of gender detection models. These advanced architectures can adeptly learn hierarchical representations of Bengali voice features, potentially uncovering subtle patterns and correlations that elude traditional machine learning algorithms.

Furthermore, the application of gender detection technology can extend beyond research settings to practical implementations in everyday contexts. Developing browser extensions or mobile applications equipped with real-time gender detection capabilities enables seamless integration into digital platforms and devices, empowering users to interact with technology in more intuitive and personalized ways.

Moreover, embedding cultural and linguistic context awareness into these applications ensures sensitivity and relevance to the diverse linguistic and sociocultural landscape of Bengali speakers. By embracing a user-centric design approach and incorporating feedback from diverse user groups, researchers can refine and tailor these applications to meet the specific needs and preferences of Bengali-speaking communities.

Additionally, exploring interdisciplinary collaborations with experts in linguistics, psychology, and gender studies offers valuable opportunities to enrich the understanding of gender identity and expression in Bengali speech.

By integrating insights from diverse disciplines, researchers can uncover deeper insights into the sociolinguistic factors shaping gender dynamics within the Bengali-speaking population. Longitudinal studies tracking changes in voice characteristics over time and across different life stages provide a longitudinal perspective on gender identity development and expression, shedding light on the dynamic nature of gender within cultural contexts. In conclusion, by embracing these multifaceted avenues of research and innovation, we can propel the field of gender detection technology forward, advancing our understanding of gender representation in Bengali voice and pioneering inclusive, ethically grounded solutions that resonate with the diverse linguistic and cultural tapestry of the Bengali-speaking community.

7.3 Limitations

One of the primary limitations of this study is the potential lack of diversity and representativeness in the Bengali voice data. If the dataset does not encompass a wide range of accents, dialects, and speaking styles, the model's performance may be biased towards the specific characteristics of the training data, reducing its generalizability to the broader Bengali-speaking population. The accuracy of the gender detection algorithms may decrease in noisy environments or real-world conditions that were not adequately simulated during the training phase.

This limitation highlights the need for further research to enhance the robustness of the models in diverse and challenging acoustic environments. While algorithms like Random Forest and Gradient Boosting demonstrate high accuracy, they also have high computational complexity. This may limit their applicability in real-time applications or on devices with limited computational resources, such as smartphones or embedded systems. The study may not fully capture the cultural nuances and variations in gender expression within the Bengali-speaking community. Further research is needed to explore these aspects comprehensively and ensure that the models respect and accurately reflect cultural diversity.

REFERENCES:

[1] Badhon, S.S.I., Rahaman, M.H. and Rupon, F.R., 2019, November. A machine learning approach to automating Bengali voice based gender classification. In 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART) (pp. 55-61). IEEE.

[2] Raahul, A., Sapthagiri, R., Pankaj, K. and Vijayarajan, V., 2017, November. Voice based gender classification using machine learning. In IOP Conference Series: Materials Science and Engineering (Vol. 263, No. 4, p. 042083). IOP Publishing.

[3] Uddin, M.A., Hossain, M.S., Pathan, R.K. and Biswas, M., 2020, August. Gender recognition from human voice using multi-layer architecture. In 2020 International conference on innovations in intelligent systems and applications (INISTA) (pp. 1-7). IEEE.

[4] Pramit, G., Goel, S. and Purwar, A., 2018. A stacked technique for gender recognition through voice. In 2018 Eleventh International Conference on Contemporary Computing (IC3).

[5] Sharma, G. and Mala, S., 2020, January. Framework for gender recognition using voice. In 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 32-37). IEEE.

[6] Jasuja, L., Rasool, A. and Hajela, G., 2020, September. Voice Gender Recognizer Recognition of Gender from Voice using Deep Neural Networks. In 2020 International Conference on Smart Electronics and Communication (ICOSEC) (pp. 319-324). IEEE.

[7] Reddy, M.M., Noushad, M.D., Goud, A.T., Bhaskar, M. and Chandu, T., 2022. GENDER VOICE RECOGNITION WITH MACHINE LEARNING. Journal of Engineering Sciences, 13(11).

[8] HaCohen-Kerner, Y. and Hagege, R., 2017. Language and gender classification of speech files using supervised machine learning methods. Cybernetics and Systems, 48(6-7), pp.510-535.

[9] Sahney, S., Babu, T., Narahari, K. and Karan, A., 2020. Identifying the Gender of a voice using Acoustic Propertie. International Journal of Advances in Engineering Research, 20(7).

[10] Gupta, M., Bharti, S.S. and Agarwal, S., 2016, December. Support vector machine-based gender identification using voiced speech frames. In 2016 fourth international conference on parallel, distributed and grid computing (PDGC) (pp. 737-741). IEEE.

[11] Kumari, M. and Ali, I., 2015, November. An efficient algorithm for Gender Detection using voice samples. In 2015 Communication, Control and Intelligent Systems (CCIS) (pp. 221-226). IEEE.

[12] Zaman, S.R., Sadekeen, D., Alfaz, M.A. and Shahriyar, R., 2021, July. One source to detect them all: gender, age, and emotion detection from voice. In 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC) (pp. 338-343). IEEE.

[13] Harb, H. and Chen, L., 2005. Voice-based gender identification in multimedia applications. Journal of intelligent information systems, 24, pp.179-198.

[14] TAŞPINAR, Y.S., SARITAŞ, M.M., Cinar, İ. and Koklu, M., 2020. Gender Determination Using

Voice Data. International Journal of Applied Mathematics Electronics and Computers, 8(4), pp.232-235.

[15] Chachadi, K. and Nirmala, S.R., 2022. Voice-based gender recognition using neural network. In Information and Communication Technology for Competitive Strategies (ICTCS 2020) ICT: Applications and Social Interfaces (pp. 741-749). Springer Singapore.

[16] Jadav, S., 2018, December. Voice-based gender identification using machine learning. In 2018 4th International Conference on Computing Communication and Automation (ICCCA) (pp. 1-4). IEEE.

[17] Gönen, S., BARIŞKAN, M.A., KARACAYILMAZ, G., Alhan, B., YILMAZ, E.N. and Artuner, H., 2022. Gender Detection Via Voice Using Artificial Intelligence Algorithms. Gazi Mühendislik Bilimleri Dergisi, 8(3), pp.567-575.

[18] Roy, P., Bhagath, P. and Das, P., 2020, May. Gender detection from human voice using tensor analysis. In Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL) (pp. 211-217).

[19] Kushwah, S., Singh, S.K., Vats, K. and Nemade, V., 2019. Gender identification via voice analysis. International Journal of Scientific Research in Computer Science, Engineering and Information Technology.

[20] Zvarevashe, K. and Olugbara, O.O., 2018, August. Gender voice recognition using random forest recursive feature elimination with gradient boosting machines. In 2018 International conference on advances in big data, computing and data communication systems (icABCD) (pp. 1-6). IEEE.

[21] Mavaddati, S., 2018. Voice-based Age and Gender Recognition using Training Generative Sparse Model. International Journal of Engineering, 31(9), pp.1529-1535.

[22] Gómez García, J.A., Moro Velázquez, L., Godino Llorente, J.I. and Castellanos Domínguez, G., 2015. Automatic age detection in normal and pathological voice.

[23] Mutiany, M. and Herlistiono, I.O., Gender Detection by Voice Using Deep Learning.

[24] Nair, R.R. and Vijayan, B., 2019. Voice based gender recognition. International Research Journal of Engineering and Technology, 6.

[25] Abakarim, F. and Abenaou, A., 2022, July. Voice gender recognition using acoustic features, mfccs and svm. In International Conference on Computational Science and Its Applications (pp. 634-648). Cham: Springer International Publishing.

[26] Burkhardt, F., Eckert, M., Johannsen, W. and Stegmann, J., 2010, May. A Database of Age and Gender Annotated Telephone Speech. In LREC.

[27] Rakesh, K., Dutta, S. and Shama, K., 2011. Gender Recognition using speech processing techniques in LABVIEW. International Journal of Advances in Engineering & Technology, 1(2), p.51.

[28] Childers, D., Wu, K. and Hicks, D., 1987, April. Factors in voice quality: Acoustic features related to gender. In ICASSP'87. IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 12, pp. 293-296). IEEE.

[29] Weston, P.S., Hunter, M.D., Sokhi, D.S., Wilkinson, I.D. and Woodruff, P.W., 2015. Discrimination of voice gender in the human auditory cortex. NeuroImage, 105, pp.208-214.

[30] Junger, J., Pauly, K., Bröhr, S., Birkholz, P., Neuschaefer-Rube, C., Kohler, C., Schneider, F., Derntl,

B. and Habel, U., 2013. Sex matters: Neural correlates of voice gender perception. Neuroimage, 79, pp.275-287.

Appendix A

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

Title:
GENDER DETECTION FROM BENGALI VOICE USING MACHINE LEARNING APPROACH

Student ID:201-15-14103

Student ID:201-15-13788

CO Description for FYDP

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

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

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

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

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	This project uses machine learning to figure out the gender of a speaker from their Bengali voice. It uses basic engineering ideas **(K3)** by applying machine learning techniques. The project shows expert knowledge **(K4)** by using advanced methods like support vector machines (SVM), to improve the accuracy of gender detection. These methods help deal with the complex features of speech.	**Page no:** [19-23] **Section:** [3.2] **Page no:** [18-19] **Section:** [3.1]
				This project uses machine learning to determine the gender of speakers from Bengali voice recordings. It applies engineering practice and design (K5) by	**Page no:** [30]

				conducting a series of experiments to refine the detection process. The project addresses engineering practice and technology (K6) by employing both machine learning models.These advanced techniques enhance the accuracy and reliability of gender detection, showcasing a strong foundation in engineering principles and specialized technological expertise.	**Section:** [3.2] **Page no:** [33] **Section:** [3.1]
				This project ensures K8 (Research Literature) by synthesizing insights from recent studies to advance gender detection from Bengali voice recordings using machine learning. It demonstrates a comprehensive understanding of current methodologies by applying both machine models. Through a series of experiments, the project refines the detection process, enhancing accuracy and reliability. This approach highlights the integration of engineering practice and design (K5) and the application of advanced technologies (K6), showcasing strong engineering principles and specialized expertise	**Page no:** [13-15] **Section:** [2.3, 2.4]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	This project addresses EP-2 by identifying challenges in gender detection from Bengali voice recordings, including limitations of traditional methods.Through comparative analysis, it confronts difficulties in understanding vocal feature distributions, offering insights to enhance diagnostic methodologies and improve accuracy.	**Page no:** [19-23] **Section:** [3.2]

3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	This project addresses EP-3 by rigorously comparing experimental outcomes in gender detection from Bengali voice recordings. It emphasizes Transfer Learning as the primary approach to enhance detection accuracy among various potential methods, showcasing a strategic choice based on comprehensive analysis.	**Page no:** [36-57] **Section:** [4.2]
4.	**EP4: Familiarity of Issues**	Yes	CO8	This project addresses **EP-3** by comparing experimental outcomes, highlighting our machine learning model as the chosen significant solution for voice detection among other discussed approaches.	**Page no:** [43-44] **Section:** [5.3]
5.	**EP5: Extends of application codes**	No	CO5	N/A	N/A
6.	**EP6: Extends of stakeholders involved and**	No	CO8	N/A	N/A
7.	**EP7: Interdependence**	Yes	CO5	Our project solves many small problems like data collection ,data preprocessing,model building etc. ultimately solving a big problem which is discussed in our report which ensures the attainment of **EP-7**.	**Page no:** [19-26] **Section:** [3.2, 3.3, 3.4]

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

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	Our project utilizes diverse resources such as high-performance computing infrastructure ,machine deep learning frameworks, annotated datasets, and ethical considerations to ensure systematic research and contribute to advancements in pneumonia voice detection.	**Page no:** [26] **Section:** [3.5]
2.	**EA2: Level of interaction**	No		N/A	N/A
3.	**EA3: Innovation**	No		N/A	N/A
4.	**EA4: Consequences for society and the environment**	Yes		This project contributes to society by improving detection methods, while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for data privacy.	**Page no:** [46-49] **Section:** [61.,6.2,6.3]

5.	**EA-5: Familiarity**	Yes		This project expands upon existing research by examining a novel approach in pneumonia detection through machine learning and through preliminary terminologies and a comprehensive comparative analysis, offering new insights into the field.	**Page no:** [10-13] **Section:** [2.1, 2.2,2.3]

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

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project addresses **CO4** by integrating effective project management and financial oversight, ensuring meticulous planning, resource allocation, and budget estimation for optimal resource utilization throughout the research lifecycle.	**Page no:** [7-8] **Section:** [1.6]
2	CO9	Yes	The project demonstrates adherence to ethical principles by prioritizing privacy, obtaining informed consent, and transparently documenting the research process, ensuring responsible knowledge dissemination and societal well-being through the ethical application of advanced technologies which comply with **CO9**.	**Page no:** [48-50] **Section:** [6.3,6.4]
3	CO10	Yes	We collected our dataset and build machine learning and these tasks were divided among our team members which ensures the attainments of **CO10**	**Page no:** [19-23] **Section:** [3.2]
4	CO12	Yes	The project's dedication to continuous learning **(CO12)** and adaptation within the dynamic technological landscape is reflected in its comprehensive data collection, rigorous statistical analysis, meticulous methodology development, and thorough experimental results and analysis, showcasing a commitment to staying updated and refining techniques to address modern challenges.	**Page no:** [18-40] **Section:** [3.2, 3.3, 3.4, 3.5, 4.2]

GENDER DETECTION FROM BENGALI VOICE USING MACHINE LEARNING APPROACH

ORIGINALITY REPORT

22%	17%	11%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	4%
2	Submitted to Daffodil International University Student Paper	4%
3	Submitted to University of Hertfordshire Student Paper	1%
4	Submitted to University of Wales Institute, Cardiff Student Paper	1%
5	researchoutput.csu.edu.au Internet Source	<1%
6	"Data Science and Applications", Springer Science and Business Media LLC, 2024 Publication	<1%
7	Neelam Nehra, Pardeep Sangwan, Neelu Trivedi. "Effect of Gender Variability in Speaker Recognition", 2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS), 2023	<1%