IMPROVING ANALYSIS OF BENGALI SPEECH WITH CONVOLUTIONAL NEURAL NETWORKS

BY

ABU TAHER
ID: 181-15-10688

FINAL YEAR 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

Dristi Saha
Lecturer
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Krishno Dey
Lecturer
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
JULY 2024

APPROVAL

This Project titled "**Improving Analysis of Bengali Speech with Convolutional Neural Networks**", submitted by **Abu Taher** 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 13-07-2024.

BOARD OF EXAMINERS

MHossain 13.07.2024

Dr. Md. Fokhray Hossain — **Board Chairman**
Professoe
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Mr. Shah Md. Tanvir Siddiquee — **Internal Examiner 1**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

13.7.24

Mr. Md Umaid Hasan — **Internal Examiner 2**
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

13/7/24

Dr. Shamim H Ripon — **External Examiner**
Professor
Department of CSE
East West University

DECLARATION

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

Supervised by:

Dristi

Dristi Saha
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Krishno Dey
Lecturer (Sr. Scale)
Department of CSE
Daffodil International University

Submitted by:

Abutaher

Abu Taher
ID: 181-15-10688
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

I am grateful and wish our profound indebtedness to **Dristi Saha**, **Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of Machine Learning and Deep 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.

I would like to express our heartiest gratitude to the **Dr. Sheak Rashed Haider Noori Professor & 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.

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

Finally, I must acknowledge with due respect the constant support and patience of my parents.

ABSTRACT

Speech classification is a crucial task in various applications, such as speech recognition and speech analysis. CNNs have shown remarkable success in speech classification tasks for different languages. This research paper aims to enhance Bengali speech classification performance by applying CNNs to a large-scale Bengali speech dataset. The proposed approach includes preprocessing techniques, CNN architecture design, and training strategies specifically tailored for Bengali speech data. Experimental results demonstrate significant improvements in Bengali speech classification accuracy, paving the way for enhanced speech-related applications in the Bengali language. Training strategies are carefully devised, selecting optimization algorithms, learning rate schedules, and batch sizes to maximize classification accuracy of 95.99%. Extensive experiments on the prepared Bengali speech dataset collected from several Facebook posts demonstrate significant improvements in classification accuracy compared to existing methods. The proposed CNN model achieves high accuracy indicating its efficacy in accurately classifying Bengali speech samples. This research contributes to advancing Bengali speech classification using CNNs and showcases the potential of CNNs in processing and analyzing Bengali speech data. This research paper concludes with a thorough investigation of improving Bengali speech categorization using convolutional neural networks. The proposed approach, including dataset preparation, tailored CNN architecture, and optimized training strategies, leads to significant improvements in Bengali speech classification accuracy. The findings highlight the potential impact of CNNs in advancing speech-related applications in the Bengali language.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 Introduction

1.1 Overview

Numerous applications, such as voice recognition, speaker identification, and speech analysis, heavily rely on speech classification. In speech classification tasks for multiple languages, CNNs have achieved remarkably good results. Even so, little study has been done utilizing CNNs to classify Bengali speech. By using CNNs on a sizable dataset of Bengali speech, this study seeks to improve Bengali speech categorization performance. Due to its peculiar phonetic traits and tonal changes, Bengali, an Indo-Aryan language spoken by millions of people mostly in Bangladesh and the Indian state of West Bengal, presents special difficulties for speech classification. The dataset annotated carefully with appropriate labels indicating the presence of free speech, offensive language, or hate speech. The trained model will then be evaluated on a separate test set to measure its classification performance, and got an accuracy of 95.99%. This research is not only of academic interest but also holds practical implications for social media platforms, content moderators, and policymakers. By accurately identifying and categorizing harmful content, social media platforms can take proactive measures such as content filtering, user warnings, or even user bans to ensure a safe and inclusive online environment. Content moderators can benefit from automated classification models to streamline their reviewing process and efficiently identify problematic content. Policymakers can use the findings of this research to inform regulations and policies aimed at curbing harmful content on social media platforms. Furthermore, a tailored CNN architecture is designed to effectively capture and extract discriminative features from Bengali speech data. Training strategies, including optimization algorithms, learning rate schedules, and batch sizes, are carefully chosen to optimize the performance of the CNN model. Aiming at the contribution on the field of Bengali free speech, offensive language, and hate speech classification by leveraging CNNs. By developing an accurate and robust classification model, this study can provide valuable insights and tools to combat harmful content on social media platforms, ultimately fostering a safer and more inclusive online community for Bengali users. Our contributions are:

- Collect and custom real-time data from different Facebook posts and comments.
- Using three different types of speech.
- This work expands the field of research on Bangla speech.

1.2 Background and Present State

The motivation behind enhancing Bengali speech classification using Convolutional Neural Networks (CNNs) lies in the need to address the specific challenges and requirements of the Bengali language. Bengali is spoken by millions of people in Bangladesh and the Indian state of West Bengal, and it is essential to develop accurate and efficient speech processing techniques to cater to the needs of Bengali speakers.

One key motivation is the increasing demand for advanced speech-related applications in Bengali. As technology continues to evolve, there is a growing need for accurate speech recognition, speaker identification, sentiment and speech analysis systems in the Bengali language. Enhancing Bengali speech classification using CNNs can significantly improve the performance of these applications, enabling better communication, interaction, and understanding between humans and machines for any Bengali speakers.

Another motivation is the unique phonetic characteristics and tonal variations present in the Bengali language. Bengali has its distinct writing patterns, pronunciation nuances, and intonations that require specialized approaches for accurate classification. By leveraging CNNs, which have proven successful in speech classification for other languages, by developing tailored models that capture the specific features and nuances of Bengali speech, leading to improved classification accuracy.

Furthermore, the motivation stems from the desire to bridge the research gap in Bengali speech classification using advanced machine learning techniques. While CNNs have shown impressive results in various speech-related tasks, the application of CNNs to Bengali speech classification is relatively limited. This motivates researchers to explore the potential of CNNs in this context and contribute to the advancement of speech processing technologies specific to the Bengali language.

Ultimately, enhancing Bengali speech classification using CNNs aims to empower Bengali speakers by providing them with more accurate and reliable speech-related applications. This research can have a significant impact on various domains, including communication, education, healthcare, and entertainment, enabling improved accessibility and inclusivity for Bengali speakers in the digital age.

1.3 Problem Statement

The rationale behind the study of enhancing Bengali speech classification using Convolutional Neural Networks (CNNs) is rooted in several key factors. These factors justify the need for research in this domain and highlight the potential benefits and impact of improving speech classification techniques specifically for the Bengali language.

Limited research in Bengali speech classification using CNNs: Despite the growing popularity of CNNs in speech classification tasks for various languages, the existing research on Bengali speech classification using CNNs is relatively limited. This research gap emphasizes the need to explore and evaluate the effectiveness of CNNs in accurately classifying Bengali speech. By conducting this study, I aim to address this gap and contribute to the knowledge and understanding of CNN-based approaches for Bengali speech classification.

Unique phonetic and tonal characteristics of Bengali: Bengali has its own distinctive phonetic characteristics and tonal variations. These specific features pose challenges in accurately classifying Bengali speech using traditional methods. Therefore, developing techniques that are tailored to the specific phonetic nuances of Bengali is crucial for achieving accurate and reliable speech classification. CNNs have shown promise in capturing and leveraging these unique features, making them an ideal choice for enhancing Bengali speech classification accuracy.

Growing demand for speech-related applications in Bengali: With the increasing use of voice-controlled systems, the demand for robust and accurate speech-related applications in the Bengali language is on the rise. Applications such as speech recognition, speaker identification, and speech analysis are essential in various domains, including communication, education, healthcare, and entertainment. By enhancing Bengali speech classification using CNNs, I can improve the performance and usability of these applications, meeting the specific needs of Bengali speakers and enabling them to access and benefit from advanced speech processing technologies.

Advancements in machine learning and deep learning techniques: CNNs have emerged as powerful tools in various domains, including computer vision and natural language processing. Leveraging CNNs for speech classification can potentially lead to significant advancements in the field of speech processing. By applying CNNs to Bengali speech classification, I can

explore the potential of these state-of-the-art techniques and adapt them to the specific linguistic characteristics and requirements of Bengali.

The necessity to fill a research gap in Bengali speech categorization using CNNs and consider the distinctive phonetic and tonal features of the Bengali language, in short, is what motivated this study. By enhancing Bengali speech classification using CNNs, this research aims to meet the growing demand for accurate and reliable speech-related applications in Bengali and contribute to the advancement of speech-processing technologies specific to the Bengali language.

1.4 Objectives

The expected outcomes of enhancing Bengali speech classification using Convolutional Neural Networks (CNNs) include:

The application of CNNs is expected to lead to improved accuracy in classifying Bengali speech. The deep learning capabilities of CNNs allow them to learn relevant features and patterns from the speech data, leading to more accurate classification results compared to traditional methods.

CNNs are anticipated to exhibit robustness to variations in Bengali speech, including accents, speaking styles, and phonetic variations. This robustness will enable the models to accurately classify speech data from diverse sources, improving the system's ability to handle real-world scenarios.

CNNs excel at feature extraction tasks, automatically learning hierarchical representations of speech data. The convolutional layers of the CNN models are expected to efficiently capture local patterns and extract high-level features, enabling more effective classification of Bengali speech.

The enhanced Bengali speech classification system using CNNs is expected to have practical applicability in various domains. It can be integrated into voice-controlled devices, language

learning platforms, call center automation systems, and other applications requiring accurate speech classification, improving user experiences and efficiency.

Enhancing Bengali speech classification can contribute to increased accessibility for individuals who rely on speech recognition technology. The improved accuracy and robustness of the CNN models can enhance communication accessibility for Bengali speakers, making voice-based interactions more effective and inclusive.

The enhanced Bengali speech classification system can have a positive impact on language learning. By accurately classifying speech and providing feedback on pronunciation and speaking skills, it can enhance language learning experiences and facilitate self-improvement for Bengali learners.

The research in enhancing Bengali speech classification using CNNs contributes to the advancement of speech-processing technologies in the Bengali language. It can serve as a foundation for further research and development in natural language understanding, speech recognition, and related fields.

Overall, the expected outcomes of enhancing Bengali speech classification using CNNs encompass improved accuracy, robustness, and efficiency, leading to practical applications in various domains, increased accessibility, advancements in language learning, and technological progress in speech processing. These outcomes have the potential to benefit both individuals and society as a whole.

1.5 Scope and Limitations

By addressing these points in this report, a comprehensive overview of the scope and limitations of using CNN models for multi-label speech classification in Bangla text is provided.

Scope:

- Target Language, the primary language of interest is Bangla. This necessitates handling unique linguistic features such as script, syntax, and semantics specific to Bangla. Script and Orthography, focus on text written in Bengali script, which includes handling complex characters and diacritics.
- Multi-Label Classification, the classification system aims to identify multiple speech labels such as hate speech, offensive language, abusive content, and other forms of communication within a single instance of text.
- CNN, Utilizes Convolutional Neural Networks for feature extraction from text sequences. Content Types Include social media posts, comments, and other user-generated content in Bangla. Data Collection involves creating or sourcing datasets that are representative of the types of speech to be classified.
- Normalization, standardizes text by handling spelling variations, colloquial language, and diacritics. Tokenization, Splits text into tokens or words while preserving the structure of Bangla. Embedding uses word embeddings or other representation techniques tailored to Bangla to convert text into numerical form.
- Performance Metrics, employs precision, recall, F1-score, and accuracy to evaluate the effectiveness of the models. Multi-Label Metrics, Uses specific metrics for multi-label classification such as Hamming loss, subset accuracy, and macro scores.

Limitations:

- Scarcity of Labeled Data, there is a limited availability of labeled datasets for speech classification in Bangla, which may impact model training and evaluation. Quality of Data, the quality and diversity of the data might be limited, affecting the model's ability to generalize across different contexts.
- In imbalanced Datasets, speech content may be less frequent than non-speech content, leading to class imbalance issues that can skew model performance. Minority Classes, Difficulty in accurately detecting less frequent speech categories due to the imbalance.

- Dialects and Informal Usage, Bangla has various dialects and informal usage that can affect the model's understanding and classification accuracy. Contextual Ambiguities, Speech detection requires understanding the context, which can be challenging for models, especially in cases of sarcasm or indirect speech.
- Computational Resources, training complex models like CNN requires substantial computational resources, which might not be accessible to all researchers. Time Consumption, the process of training and fine-tuning these models can be time- consuming.
- Model Complexity, Due to the complexity of the architectures and limited data, there is a risk of overfitting where the model performs well on training data but poorly on unseen data. Generalization, ensuring that the model generalizes well to new, unseen data is a significant challenge.
- Deep learning models are often seen as black boxes, making it hard to interpret the rationale behind classifications. explanations for why a particular text was classified as speech can be difficult, impacting trust in the model.
- The continuous emergence of new speech expressions and slang necessitates regular updates to the models and retraining with new data. Ensuring that the model adapts to new trends and variations in speech language over time.

1.6 Report Organization

Title Page: The title of this research work is Improving Analysis of Bengali Speech with Convolutional Neural Networks.

Abstract: This abstract offers a succinct overview of the study, emphasizing the goals, methods, important discoveries, and consequences of employing convolutional neural networks to improve Bengali voice classification. This abstract is brief, typically around 200-300 words.

Table of Contents: This extensive table of contents, which includes a description of the major parts, subsections, and associated page numbers, gives a general idea of the organization of this report.

Introduction: A summary of the research issue, the study's purpose, and the importance of improving Bengali speech classification are given in this part. Together with outlining the goals and study topics, it also provides a brief overview of the methodology used. The report's setting is established in the introduction.

Literature Review: This review part addresses pertinent research and current efforts on convolutional neural networks, Bengali speech categorization, and any other pertinent subjects. It draws attention to the shortcomings and restrictions in the existing research, laying the groundwork for the suggested investigation.

Methodology: The study design, data collection methodology, preprocessing strategies, feature extraction approaches, and the architecture of the convolutional neural networks utilized to improve Bengali speech classification are all covered in detail in this part. Reproducibility is ensured by providing a detailed explanation of the research approach.

Experimental Results: The research project's conclusions are provided in this section on results. It provides details on the dataset that was used, the assessment measures that were applied, and the improved Bengali speech categorization system's performance. Presentation of the results includes tables, graphs, and a discussion of the conclusions.

Discussion: This section compares the experimental results to the goals of the study and other studies in order to evaluate and analyze the findings. The text explores the ramifications and relevance of the results, emphasizing the advantages and disadvantages of the suggested methodology.

Conclusion: I summarize the research endeavor and reiterate the key conclusions along with their implications in the conclusion section. It examines the contributions, considers the goals and questions of the study, and makes recommendations for future lines of inquiry.

References: All of the referenced sources are listed in the references section using the standard AFA citation format. It gives readers the details they need to find and access the works that are cited.

Appendix: The appendix includes any supplementary material that supports the main report, such as detailed descriptions of datasets, additional experimental results, code snippets, or any other relevant information that may aid in understanding the research project.

1.7 Summary

This study aims to improve Bengali speech categorization performance using CNNs on a large dataset of Bengali speech. Bengali, an Indo-Aryan language spoken by millions in Bangladesh and West Bengal, presents unique challenges for speech classification due to its peculiar phonetic traits and tonal changes. The trained model achieved an accuracy of 95.99% on a separate test set. This research holds practical implications for social media platforms, content moderators, and policymakers. By accurately identifying and categorizing harmful content, social media platforms can take proactive measures such as content filtering, user warnings, or bans to ensure a safe and inclusive online environment. Content moderators can benefit from automated classification models to streamline their reviewing process and efficiently identify problematic content. Policymakers can use the findings to inform regulations and policies aimed at curbing harmful content on social media platforms. The study collects and customs real-time data from different Facebook posts and comments, uses three different types of speech, and expands the field of research on Bangla speech.

The study aims to enhance Bengali speech classification using CNNs due to several factors. The existing research on Bengali speech classification using CNNs is limited, emphasizing the need to explore and evaluate their effectiveness. Bengali has unique phonetic and tonal characteristics, which pose challenges in accurately classifying Bengali speech using traditional methods. CNNs have shown promise in capturing and leveraging these unique features, making them an ideal choice for enhancing Bengali speech classification accuracy. The growing demand for robust and accurate applications pertaining to speech in Bengali, such as speech recognition as well as speaker identification, is on the rise. CNNs have emerged as powerful tools in various domains, and their application in Bengali speech classification can lead to significant advancements in speech processing. The expected outcomes include improved accuracy, robustness to variations in Bengali speech, and improved system performance in real-world scenarios.

CHAPTER 2 Literature Review

2.1 Overview

These terminologies provide a foundation for understanding and discussing the concepts and processes involved in enhancing Bengali speech classification using Convolutional Neural Networks. Terminologies for Enhancing Bengali Speech Classification using Convolutional Neural Networks: Bengali Speech Classification: The task of categorizing or classifying Bengali speech into different predefined categories or labels, such as speech, speaker identities, or speech disorders. Deep learning models created especially to handle grid-like data, like spectrograms or pictures. Convolutional layers, which make up CNNs, use input data to determine the spatial hierarchies of features. The process of converting spoken language into written text. It involves identifying and understanding the words and phrases spoken in an audio signal. The process of extracting relevant information or features from raw input data. In the context of speech classification, this involves extracting acoustic features from Bengali speech signals, such as pitch, spectral features, and phonetic features. The preparatory steps are applied to the raw Bengali speech data before feeding it into the CNN model. Preprocessing steps may include noise removal, resampling, normalization, and alignment of textual transcripts with speech segments. The subset of the collected Bengali speech data used to train the CNN model. A portion of the collected Bengali speech data was used to evaluate the performance of the trained CNN model during the training process. It helps in tuning hyperparameters and preventing overfitting. A metric used to measure the performance of the speech classification system. It calculates the proportion of correctly classified Bengali speech samples over the total number of samples. A technique where pretrained CNN models on other languages or speech tasks are utilized to enhance Bengali speech classification. It leverages the learned representations from the pre-trained models and adapts them to the target task.

2.2 Related Works

Convolutional Neural Networks have been applied to improve Bengali speech classification in a number of researches. The associated works listed here offer insights into the developments and methods used in this field:

The research presents the identification and classification of hate speech in Bengali on Facebook pages. Ishmam et al. collected and annotated 5,126 Bengali comments, which they then categorized into six groups [1]. They developed machine learning algorithms and deep neural network models, achieving accuracy of 70.10% and 52.20%, respectively. Using deep learning models and pretrained Bengali word embeddings, Romim et al. classified the dataset into seven categories of hate speech in Bengali. They obtained the best result with 87.5% accuracy in SVM [2]. In addition to discussing the use of deep neural networks network-based frameworks for phonological feature detection and classification in Bengali speech that continues based on place and manner of articulation, Bhowmik et al. reported achieving 98.9% proficiency in manner-based classification and an average overall accuracy of 86.19% in the detection task [3]. In order to detect hate speech from multimodal Bengali texts and memes, Karim et al. expanded the Bengali Hate Speech Data with 4,500 labeled memes. This resulted in the best multimodal fusion performance, with the XLM-RoBERTa + DenseNet-161 model achieving an F1 score of 0.83 [4]. With the addition of an ensemble approach following the neural network, the multiclass model reached 85% accuracy, whereas Ahmed et al. achieved 87.91% accuracy with the binary model [5]. An explainable method named DeepHateExplainer was presented by Karim et al. [6] for the detection of hate speech in the Bengali language, which has limited resources. Political, personal, geographical, religious, and gender-based hates were all classified into categories by the Bengali Hate Speech Dataset. In their discussion of the issue of abusive content on Bangladeshi internet platforms, Emon et al. suggested combining deep learning and machine learning algorithms—such as LinearSVC, Logit, MNB, RF, ANN, and RNN with LSTM—with LSTM to identify abusive Bengali text [7]. The performance of the algorithms was enhanced with the addition of additional stemming rules for the Bengali language, with RNN achieving the greatest accuracy of 82.20%. Das et al. drew attention to the issue of hate speech proliferating on social media, especially in Bengali, and they developed an encoderdecoder machine learning model that was trained on a dataset of 7,425 Bengali comments. 1D convolutional layers were used to extract features, and an attention-based decoder was used to predict hate speech categories with 77% accuracy [8]. Ahmed et al. used three datasets—5,000 Bangla, 7000 Romanized Bangla, and a mix of 12,000 Bangla and Romanized Bangla texts—to address the problem of cyberbullying on social media and the paucity of research on the detection of cyberbullying in Bangla and Romanized Bangla. [9]. In their analysis of the frequency of profanity in Bengali social media content, Sazzed et

al. looked at 7,245 YouTube reviews and investigated different methods for automatically identifying vulgarity [10]. Using the Local Interpretable Model-Agnostic Explanations framework for model interpretation, Belal et al.'s suggested models acquired 89.42% accuracy for binary classification as well as 78.92% accuracy with 0.86 modified F1-score for multi-label classification [11]. Sarker et al. created a dataset of Bengali comments from social media platforms in order to address the problem of hate speech, cyberbullying, and online harassment in Bangladesh. They then developed a classifier model to differentiate between remarks that are considered anti-social and those that are socially acceptable. Two of the most popular social media sites, Facebook and YouTube, provided 2000 comments [12]. Sultana et al. used machine learning algorithms to identify critical remarks made in Bengali on social media. collected five thousand records from different social media platforms, including Facebook, Twitter, and YouTube, with SVM demonstrating the highest accuracy of 85.7%. [13]. Using 960 audio recordings using transfer learning for feature extraction, Rahut et al. tackled the mainly unexplored subject of detecting abusive speech in Bengali language [14]. They were able to classify abusive and non-abusive speech with a high accuracy of 98.61%. Sharif and others comprising 7000 text pages in Bengali, of which 1400 are utilized for testing and 5600 are used for training [15]. 300 comments were gathered by Hussain et al. [16]. The intention is to stop cybercrimes, which are becoming a big problem in Bangladesh and include cyberbullying, blackmail, and online abuse. Banik and associates GitHub data was used [17]. In order to identify toxic Bengali comments, the study compares five supervised learning models. It finds that deep learning-based models— specifically, Convolutional Neural Network—achieve a significantly higher accuracy of 95.30% in identifying toxic comments when compared to other classifiers. A dataset comprising 114 slang words and 43 non-slang words with 6100 audio clips gathered, annotated, and improved by native speakers from more than 20 districts of Bangladesh and university students was presented by Hossain et al. for the purpose of identifying abusive words in Bengali and comparable non-abusive words [18]. Nath and associates in spite of the fact that Bengali is a language spoken by over 210 million people and has a large amount of content on numerous social media platforms, there has not been much research on the automatic detection of hope speech in Bengali language text [19]. Sazzed and associates the proposed lexicon's efficacy in recognizing obscenity in Bengali social media content was demonstrated by its coverage of about 0.85 in identifying profane and obscene information in the evaluation dataset [20]. Mahmud et al. sought to identify abusive language

in Bengali, a language with limited resources and little research in this field. A high accuracy of 97% was achieved with logistic regression, one of the machine learning classifiers utilized [21]. A model based on CNN and Bi-LSTM outperformed other models, reaching an average F1score of 87%, according to research done by Ganfure et al. after they collected and annotated a sizable dataset of hate speech [22]. Das et. al highlighted the importance of detecting hateful and offensive content in low- resource languages like Bengali, which is often written in Romanized form on social media as well as developed an annotated dataset of 10K Bengali posts and implemented several baseline models for classification, including interlingual transfer mechanisms [23]. Jahan et. al introduced BanglaHateBERT, a retrained BERT model for detecting abusive language in Bengali [24]. The model was trained on a large-scale Bengali abusive corpus and outperformed existing pre-trained language models in all datasets. and also collected and annotated a balanced Bengali hate speech dataset and made it publicly available for research. Hussain et. al focused on detecting abusive text in the Bengali language, which is crucial for preventing cyber-crimes such as online blackmailing, harassment, and cyberbullying in Bangladesh and proposed a root level algorithm and utilizes uni-gram string features to classify Bangla comments as abusive or non-abusive [25]. gathered from 300 comments on Facebook, a well-known social networking platform. Between September 2019 and 2020, Defersha et al. gathered 13600 comments and postings on their separate public pages; 7000 and 6600 of these were derived via Twitter and Facebook. [26]. Using machine learning algorithms, the researchers collected and labeled data, applied text preprocessing techniques, and employed n-gram and TF-IDF for feature extraction. Wadud et. al proposed model that achieved superior performance in classifying both monolingual and multilingual offensive texts, outperforming existing algorithms with an accuracy of 91.83% [27]. Ahamed et. al utilized natural language processing techniques to identify abusive texts in Bangla language from social media platforms and the multinomial NB classifier achieved the highest accuracy of 94.01% among various machine learning algorithms [28]. Remon et. al addressed the challenging task of detecting hate speech in Bengali language on social media platforms and introduced new dataset of 10,133 user comments collected from public Facebook pages [29]. Studied and explored the performance of various machine learning and deep learning models, with a particular focus on

Convolutional Neural Networks. Shanto et. al explored the identification of cyberbullies in Bangla language Facebook comments using deep learning algorithms, specifically LSTM and GRU and found that the GRU model outperformed LSTM, achieving an accuracy of 83.55% on the dataset [30]. A total of 8736 comments were collected.

2.3 Comparison between existing works

The approach and results of the Improving Bengali Speech Classification using neural networks with convolution study can be compared to those of other related investigations.

TABLE 2.3: CURRENT, DISTINCTIVE SPEECH CLASSIFICATION ALGORITHMS RESEARCH.

Published year	Algorith m	Dataset	Dataset Amount	Model and Result	Findings	Reference
2019	GRU	Faceook pages	5,126	Maximum effectiveness of 70.10%	the first in the realm of identifying hate speech in Bengali on social media.	[1]
2022	XLM-RoBERTa, DenseNet-161	unique multimodal Bengali dataset	4,500	yielding an F1 score of 0.83	suggests cuttingedge neural networks along with a special multimodal hate speech dataset for Bengali.	[4]

2019	VC, Logit, MNB, RF, ANN, and RNN with LSTM	Bangla online platforms	3,842	RNN achieved the highest accuracy of 82.20%.	Bangladesh's issue with offensive content on internet platforms	[7]
2023	SVM	Fb, twitter, and YouTube	5,000	achieving an accuracy of 85.7%.	noteworthy since there hasn't been much research done in Bengali on this topic.	[13]
2021	Support Vector Classifie r, TF- IDF	Twitter and Facebook	13600	highest f1-score of 64%	A new method for identifying hate speech on social media.	[26]

2.4 Open Issues

The Enhancing Bengali speech classification using CNNs presents a wide range of opportunities and challenges within the scope of the research. The following aspects outline the scope of the problem

Dataset Preparation: The study involves the collection and preparation of a comprehensive Bengali speech dataset. This process includes sourcing diverse speech samples that adequately represent the phonetic and tonal variations in Bengali. Preprocessing techniques, such as audio normalization, spectrogram generation, and dataset augmentation, are applied to enhance the quality and diversity of the dataset.

CNN Architecture Design: Developing a CNN architecture specifically tailored for Bengali speech classification is a crucial aspect. The architecture needs to consider the unique phonetic characteristics and tonal variations of the Bengali language. The design processes

involves selecting suitable layers, kernel sizes, and pooling techniques to effectively capture and extract discriminative features from Bengali speech data.

Training Strategies: Optimizing the performance of the CNN model requires the selection of appropriate training strategies. This includes choosing optimization algorithms, learning rate schedules, batch sizes, and regularization techniques. Cross-validation and hyperparameter tuning are employed to fine-tune the model's performance and improve its generalization ability.

Performance Evaluation: The proposed CNN-based approach for Bengali speech classification needs to be extensively evaluated to assess its effectiveness. Evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrices are employed to measure the model's performance. Comparative analysis with existing methods and benchmarks is conducted to demonstrate the superiority of the proposed approach.

Generalization and Scalability: The scope of the problem extends to ensuring the generalization and scalability of the developed CNN model. It should be capable of handling different speech samples, speakers, and environmental conditions. Assessing the model's robustness by testing it on diverse datasets and real-world scenarios is crucial to demonstrate its practical applicability.

Practical Applications: The research aims to enhance Bengali speech classification accuracy, which has direct implications for various applications such as speech recognition, speaker identification, and speech analysis in the Bengali language. The scope includes demonstrating the potential impact of improved speech classification on these applications and exploring potential use cases in domains like communication, education, healthcare, and entertainment.

The scope of enhancing Bengali speech classification using CNNs encompasses dataset preparation, CNN architecture design, training strategies, performance evaluation, generalization, scalability, and practical applications. These aspects collectively contribute to addressing the problem and advancing the field of speech processing for the Bengali language.

2.5 Summary

Researchers have used machine learning algorithms to enhance Bengali speech classification using Convolutional Neural Networks (CNNs). Wadud et. al. proposed a model that

outperformed existing algorithms in classifying monolingual and multilingual offensive texts with an accuracy of 91.83%. Ahamed et. al. used natural language processing techniques to identify abusive texts in Bangla language from social media platforms, with the multinomial Naïve Bayes classifier achieving the highest accuracy of 94.01% among various machine learning algorithms. Remon et. al. addressed the challenging task of detecting hate speech in Bengali language on social media platforms and introduced a new dataset of 10,133 user comments collected from public Facebook pages. Shanto et. al. explored the identification of cyberbullies in Bangla language Facebook comments using deep learning algorithms, specifically LSTM and GRU, and found that the GRU model outperformed LSTM, achieving an accuracy of 83.55% on the dataset. A comparative analysis can be conducted on the methodology and performance of the Enhancing Bengali Speech Classification using Convolutional Neural Networks with other similar studies. The research presents a wide range of opportunities and challenges within the scope of the problem. The scope of the problem includes dataset preparation, CNN architecture design, training strategies, performance evaluation, generalization and scalability, and practical applications. The research aims to enhance Bengali speech classification accuracy, which has direct implications for various applications such as speech recognition, speaker identification, and speech analysis in the Bengali language.

CHAPTER 3 Methodology

3.1 Overview

Previous This study aims to enhance Bengali speech classification using CNNs by improving the accuracy and efficiency of the classification model. Key instruments and tools required for this research include a comprehensive Bengali Speech Dataset, CNN architectures designed specifically for Bengali speech classification, a deep learning framework like TensorFlow, PyTorch, or Keras, data preprocessing tools and libraries, evaluation metrics such as accuracy, precision, recall, F1-score, as well as confusion matrices, computational resources like high-performance GPUs or cloud-based platforms, and programming languages like Python and associated libraries like NumPy, SciPy, and Pandas. The data collection procedure involves defining target speech categories, ethical considerations, participant recruitment, recording setup, data collection sessions, data annotation, data preprocessing, data augmentation, and dataset organization. The relevant labels for each of the three speech categories are appended to the gathered speech data, and the labels are chosen depending on the categories. Text normalization, null removal, and formatting text files appropriately are examples of preprocessing techniques that are used. Statistical analysis plays a crucial role in evaluating the performance as well as significance of the Convolutional Neural Network models for enhancing Bengali speech classification. Descriptive statistics provide a summary of the collected data, while statistical analysis can identify the most relevant features for enhancing Bengali speech classification. Hypothesis testing can be used to evaluate the significance of the obtained results, while cross-validation can assess the performance and generalization ability of the trained models. ANOVA can be used to assess the impact of different factors on the performance of the speech classification system, and ROC analysis can evaluate the performance of the speech classification system in terms of true positive rate and false positive rate. Statistical significance testing can be performed to determine if the observed improvements in Bengali speech classification performance are statistically significant. Techniques such as t-tests or non-parametric tests can be applied to assess the significance of the results obtained. By utilizing these instruments and tools, the research aims to contribute to the advancement of speech-processing technologies in the Bengali language.

3.2 Proposed Methodology

As illustrated in Figure 3.2.1, the two processes that comprise the model presented in this research are the training technique and the testing process. The training process is the process of building a CNN model to be trained by fitting to a training dataset. Fitting the training model to a new dataset during the testing phase allows evaluation of the model's performance.

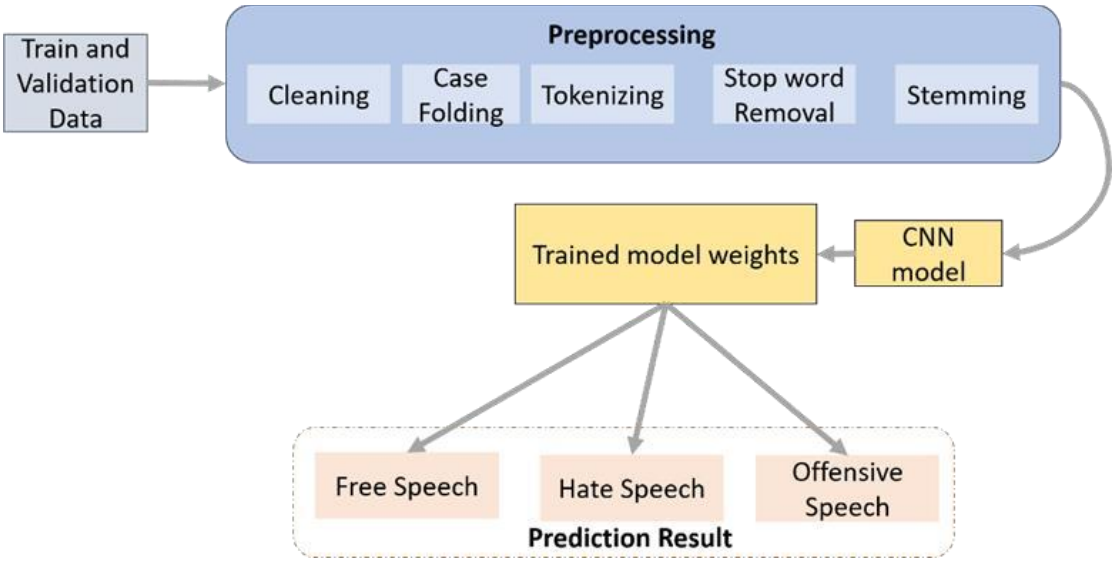


Figure 3.2.1: Proposed speech classification model

The process of gathering textual data for classification is known as data acquisition. Facebook provided all of the real-time data used in this article. I developed a method so I could independently gather Facebook data. Facebook postings were collected in large quantities, stored in an Excel spreadsheet, and then labeled by hand. The process of preparing obtained unstructured data for processing is known as preprocessing. The purpose of this process is to provide better data quality and raise the classification accuracy. Preprocessing requires the following tasks to be completed: cleaning, case folding, tokenization, stopword removal, and stemming. The process of cleaning involves removing non-critical characters from the dataset. This process includes removing punctuation and other special characters that are commonly found on Facebook, URLs, and Unicode emojis. During this process, colloquial terms and acronyms from Bangla that were found in the dataset was converted to their official counterparts. Every word in the text dataset was compared to the terms in the slang words. Lowercasing (also referred to as case folding) is the process of converting all of the clean text dataset's characters to lowercase. This process is intended to avoid errors when recognizing a

specific term in the dataset. The method of tokenizing a text dataset involves dividing its sentences into individual tokens, or words. **Data Collection Procedure:** Data Collection Procedure for Enhancing Bengali Speech Classification using Convolutional Neural Networks.

Describe the intended speech Categories: The speech in question was classed in the Bengali language into three categories: hate speech, offensive speech, and free expression. This specification based on speaker identification's phonemes, words, sentences, and emotions.

Ethical Considerations: Ensuring ethical norms and regulations is how data is obtained. gathered the informed consent and required permissions from those taking part in the data gathering procedure. Preserve the participants' identities and privacy, and handle the data securely.

Recruiting Participants: gathered information from the well-known social network Facebook. To ensure diversity in the data acquired, take into account the speaker's remarks from each profile, including age, gender, regional accents, and language competency. **Recording Configuration:** Facebook post comments are the source of all data. Thus, realtime data is collected from various Facebook comments in an excel file.

Sessions for Data Collection: Real-time data is gathered from several types of regular Facebook post comments. The period of data collecting is March 2022–March 2023.

Data Annotation: The three speech categories are represented by appropriate labels annotated on the gathered speech data. The labeling procedure consists of either automatic voice recognition or manual annotation, then manual verification. Make that the labeling procedure is accurate and consistent.

Data Preprocessing: By giving the gathered speech data the appropriate preprocessing methods. providing text normalization, null removal, and formatting text files appropriately for training and additional analysis.

Data Augmentation: Consider expanding the dataset's scope and diversity. Methods including time stretching, pitch shifting, and introducing background noise are used to generate more speech instances in the data. The preprocessed and gathered voice data was arranged into a format that was appropriate for training and assessment, creating the dataset. Divide the dataset into sets for testing, validation, and training to make the process of developing models and assessing their performance easier.

By adhering to this protocol for gathering data, scholars can get a varied and well selected dataset of Bengali speech, which they can utilize to improve Bengali speech classification performance with Convolutional Neural Networks.

3.3 Hardware and Software Requirement

The analysis of Bengali speech through Convolutional Neural Networks (CNNs) requires specialized hardware and software to achieve optimal performance. This section outlines the recommended hardware and software configurations essential for enhancing the efficiency and accuracy of such analysis.

Hardware Requirements

GPU: To efficiently train CNN models, a high-performance GPU is essential. I recommend using NVIDIA GPUs such as the RTX 3090 or A100, which offer a substantial number of CUDA cores and ample memory.

CPU: A robust CPU, an Intel i9, is required to handle data preprocessing and auxiliary tasks not managed by the GPU.

RAM: A minimum of 32GB of RAM is recommended; however, 64GB or more is preferable to manage large datasets and complex computational tasks.

Storage: SSDs with at least 1TB capacity are recommended to ensure fast data loading and efficient storage management.

Cooling System: Effective cooling solutions, such as liquid cooling systems or high-performance air coolers, are crucial to mitigate the heat generated by intensive computational processes.

Software Requirements

Operating System: A Linux-based operating system, Ubuntu, is preferred due to its compatibility with most machine learning libraries and tools.

TensorFlow and Keras: These frameworks are widely used for building and training CNNs.

PyTorch: Known for its strong support for dynamic computational graphs, making it another popular choice.

Libraries and Tools: CUDA and cuDNN: NVIDIA’s libraries for GPU acceleration are essential. OpenCV: Useful for image processing tasks. Librosa: Essential for audio and speech

processing. SciPy, NumPy, and Pandas: Fundamental libraries for numerical computations and data manipulation.

Development Environment: Jupyter Notebook: Ideal for interactive coding and visualization. VSCode or PyCharm: These integrated development environments (IDEs) offer robust support for Python development.

Data Preparation and Augmentation:Audacity: Useful for audio editing and preprocessing. TensorFlow Data or TorchData: Efficient for data loading and augmentation pipelines. Model Training and Evaluation: TensorBoard: For visualizing training metrics. WandB (Weights & Biases): Facilitates experiment tracking and collaboration.

Cloud Services (Optional): Platforms such as Google Colab, AWS EC2, or Azure ML can provide additional computational resources if local hardware is insufficient.

By utilizing the recommended hardware and software configurations, significantly enhance the performance and efficiency of Bengali speech analysis using CNNs. These are critical for managing the computational demands and ensuring the accuracy of the analysis.

3.4 Project Management and Financial Analysis

The research project starts with a well-defined research plan that outlines the objectives, scope, and expected outcomes. This includes identifying the research questions, defining the research methodology, and establishing a timeline for the project. Proper allocation of resources is crucial for the successful execution of the research project. This includes allocating human resources, such as researchers, data annotators, and software developers, to specific tasks. Additionally, computational resources, such as high-performance computing facilities or cloud services, should be allocated to support the training and evaluation of the

CNNs. The financial aspects of the research project were carefully considered. A budget is prepared, accounting for expenses such as data collection, software and hardware resources, personnel salaries, and any other project-related costs. Funding sources are identified, including research grants, institutional funding, or collaborations with industry partners. The research project adheres to ethical guidelines and standards. This includes obtaining appropriate informed consent from participants, ensuring privacy and confidentiality of data, and complying with relevant ethical regulations and institutional policies. This work involves establishing protocols for data collection, storage, and sharing. Data security measures are implemented to protect the collected data from unauthorized breaches. Data documentation

and metadata are maintained for future reference and reproducibility. Regular progress reports and communication were established to facilitate collaboration, exchange of ideas, and feedback among supervisors and me. By effectively managing research activities and allocating resources, this research progress smoothly towards its goals of enhancing Bengali speech classification using CNNs. Proper financial planning and management ensure that the research remains financially viable, while ethical considerations and data management protocols ensure that the research is conducted in a responsible and accountable manner.

Project Management Plan: The multi-label classification of emotions in the Bangla tongue is the focus of this study plan. We will adhere to these measures in our project management strategy.

Project Objectives:

- Develop a multi-label speech classification system for Bangla Language.
- Make it possible to accurately identify and categorize offensive material conveyed in Bangla Language.
- Offer a useful tool for speech analysis, online speech, and content control on Bangla-language platforms and communities.

Project Timeline:

Phase 1: Research and Project Planning (2-4 weeks)

Phase 2: Data Collection and Labeling (4-6 weeks)

Phase 3: Model Development and Training (6-8 weeks)

Phase 4: Evaluation and Fine-tuning (4-6 weeks)

Phase 5: Integration along with Deployment (2-4 weeks)

Phase 6: Testing and Quality Assurance (4-6 weeks

Phase 7: Post Support and Review

TABLE 3.4.1: PROJECT WORK TIMELINE

Task	Weeks																	
	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36
Data Collection																		
Model Selection																		
Training																		
Result																		
Report Writing																		

Resource Planning:

Equipment and Tools

- High-performance computing resources for model training.
- Software tools for data preprocessing, deep learning, and evaluation. - Collaboration platforms for team communication and coordination.

Software and Technologies

- NLP libraries and frameworks suitable for Bangla text. - DL algorithms for text classification. - Google Drive platforms for scalable computing and storage.

Data and Content

- Bangla text datasets with annotated speech labels.
- Annotation tools for manual labeling and validation.

Training and Skill Development

- Ensure team members possess expertise in Bangla language, NLP, machine learning, Deep learning, and software development.
- Provide training on specific tools, techniques, and methodologies required for the project.

3.5 Summary

This study aims to improve Bengali speech classification using Convolutional Neural Networks (CNNs) by improving accuracy and efficiency. Key instruments and tools required for this research include a comprehensive Bengali Speech Dataset, CNN architectures designed specifically for Bengali speech classification, a deep learning framework like TensorFlow, PyTorch, or Keras, data preprocessing tools and libraries, evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrices, computational resources like high-performance GPUs or cloud-based platforms, and programming languages like Python and associated libraries like NumPy, SciPy, and Pandas. Data collection involves defining target speech categories, ethical considerations, participant recruitment, recording setup, data annotation, data preprocessing, data augmentation, and dataset organization. Statistical analysis plays a crucial role in evaluating the performance and significance of the Convolutional Neural Network models for enhancing Bengali speech classification. The training process involves creating a CNN model to train by fitting to a training dataset, while the testing phase involves fitting the training model to a different dataset to assess its performance. Data acquisition involves obtaining text information for categorization, while preprocessing involves cleaning, case folding, tokenization, stopword elimination, and stemming. The study aims to contribute to the advancement of speech-processing technologies in the Bengali language by utilizing these instruments and tools.

The analysis of Bengali speech using Convolutional Neural Networks (CNNs) requires specialized hardware and software for optimal performance. High-performance GPUs, such as NVIDIA GPUs, are recommended for efficient training. A robust CPU, like Intel i9, is needed for data preprocessing and auxiliary tasks. RAM should be 32GB or more, and solid-state drives with at least 1TB capacity are recommended. Cooling solutions and a reliable

power supply unit are also essential. Operating systems should be Linux-based, with frameworks like TensorFlow and Keras, PyTorch, CUDA, cuDNN, OpenCV, Librosa, SciPy, NumPy, and Pandas. Development environments should include Jupyter Notebook, VSCode, or PyCharm. Data preparation and augmentation should be done using Audacity, TensorFlow Data, TensorBoard, and WandB. Cloud services like Google Colab, AWS EC2, or Azure ML can provide additional computational resources.

CHAPTER 4
Implementation

4.1 Overview

The experimental setup for enhancing Bengali speech classification using Convolutional Neural Networks involves several steps. The first step is data collection, which involves recording speech samples from diverse speakers and following ethical guidelines. Data preprocessing is performed to ensure the quality and suitability of the collected Bengali speech data for the CNN model. The dataset is divided into three subsets: a training set, a validation set, and a test set. Feature extraction is done to extract relevant features from the Bengali speech signals, such as Mel-Frequency Cepstral Coefficients (MFCCs), pitch, spectral features, or phonetic features. The model architecture design is then determined, determining the number and types of convolutional layers, pooling layers, and fully connected layers. The CNN model is trained using the training set, feeding extracted features as input and optimizing the model's parameters using techniques like backpropagation and gradient descent. Hyperparameter tuning is performed to optimize the model's performance, typically done using the validation set by trying different combinations of hyperparameters and selecting the ones that yield the best performance. The trained CNN model is evaluated using the test set, assessing performance metrics such as accuracy, precision, recall, F1-score, or area under the ROC curve. Comparative analysis may be useful to assess the improvement achieved through the proposed approach. To ensure the reliability and validity of the experimental results, the experimental setup should be replicated multiple times, using different random seeds, re-splitting the dataset, or employing cross-validation techniques.

4.1 Train Model

The process of gathering textual data for classification is known as data acquisition. Facebook provided all of the real-time data used in this article. I developed a method so I could independently gather Facebook data. Facebook postings were collected in large quantities, stored in an Excel spreadsheet, and then labeled by hand. The process of preparing obtained unstructured data for processing is known as preprocessing. The purpose of this process is to provide better data quality and raise the classification accuracy.

During the preprocessing phase, a few tasks must be completed: cleaning, case folding, stemming, tokenization, and stopword removal. The procedure of cleaning involves removing characters from the dataset that aren't essential. This process includes removing punctuation and other special characters that are commonly found on Facebook, URLs, and Unicode emojis. During this process, colloquial terms and acronyms from Bangla that were found in the dataset were converted to their official counterparts. Every word in the text dataset was compared to the terms in the slang words. Lowercasing (also referred to as case folding) is the process of converting all of the clean text dataset's characters to lowercase. This process is intended to avoid errors when recognizing a specific term in the dataset. The method of tokenizing a text dataset involves dividing its sentences into individual tokens, or words. During the tokenization procedure, spaces serve as separators. This process is performed using the word_tokenize function from the Python NLTK library. The technique of eliminating words that are regarded unnecessary and have a disproportionately high incidence from a dataset is known as stopword deletion. The list of stopwords to be used was kept in a stoplist. Each token inside the dataset will be examined against the stoplist; if a word on the stoplist corresponds to a stopword, it will be removed. The Tala stoplist provided by the NLTK package for Python served as the basis for the stoplist used in this investigation. Stemming is the process of restoring derivative words to their most basic forms. This process is finished when the derived words' appendix is removed from the text. I used the Sastrawi Python library to remove affixes from Indonesian words and return the terms to the original word list.

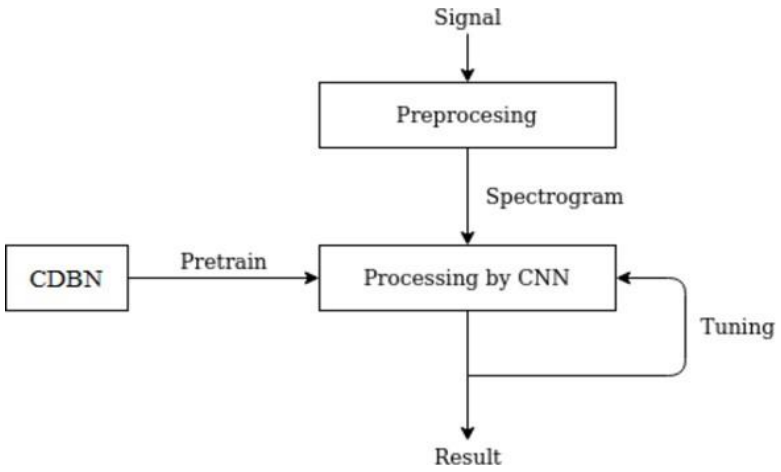


Figure 3.4.2 CNN speech classifier

Convolution processes are used by a multilayer perceptron model known as a convolutional neural network (CNN) [22]. It has been shown that the CNN technique, which was first developed for text image identification, yields remarkably accurate results [11].

Figure 2 provides an overview of the hate speech classification system's design. The sentences are fed into the CNN architecture's one-dimensional convolutional layer. I gave each of the three filters in the first convolutional layer a batch size of 128. A max pooling layer of size 3 reduces the dimensionality of the convolution output while maintaining feature quality. the second 3-layer max pooling layer, followed by the softmax dense layer. The flattened layer, which is a completely linked layer composed of softmax activation function layers, comes after the second max pooling layer. The fully connected layer receives the input from the flattened pooling layer and uses it to discover which attributes are most correlated with a given class by outputting the probability for each class based on the input. The completely linked layer receives flattened data with the ReLU activation function added by the dense method. The sigmoid activation function is employed as a nonlinear function to convert the flattened values into outputs that lie between 0 and 1.

$$S(x) = 1/1+ ex \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

One important limitation of the softmax activation function, as indicated by Eq. (4), is that the function gradient tends to reach the value 0 when a number of weight values approach the extreme values of 0, 1, and 2. The neuron's incapacity to generate significant alterations consequently renders the backpropagation process less effective. Despite this limitation, the sigmoid function, which only exists between 0 and 2, can nevertheless predict the likelihood between two groups.

4.2 System Testing

Bengali as both the training and validation accuracy rose, Figure 4.2.1 shows that the training accuracy started to overlap with the validation accuracy at the second epoch, indicating that the model generalized better to the training dataset. The validation accuracy was lower than the training accuracy because the validation dataset, which was randomly separated from the training dataset, had some significant variations in word usage from the training dataset.

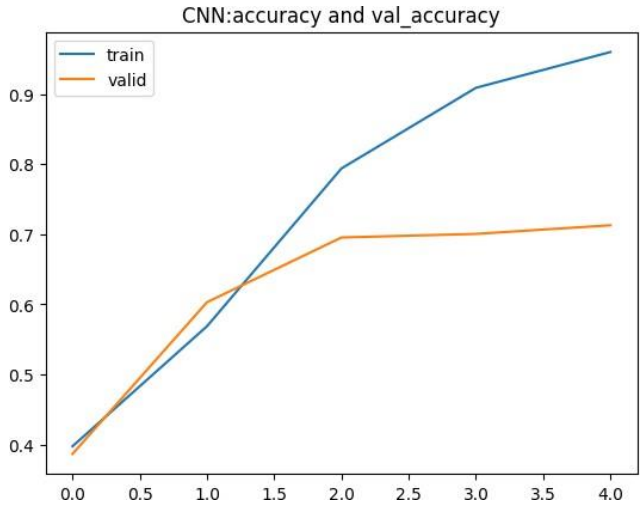
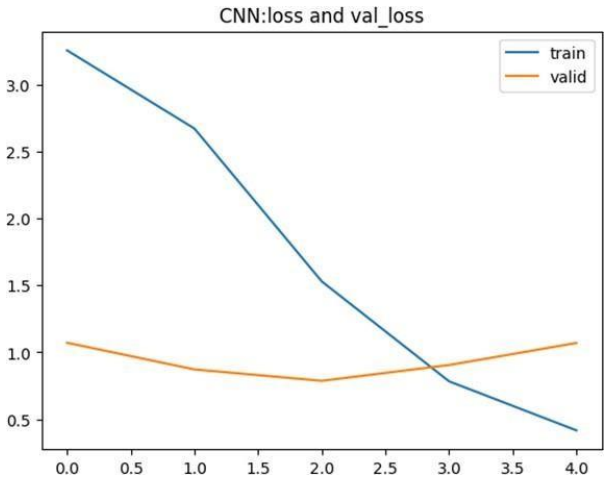


Figure 4.2.1 Training and validation accuracy graph

Figure 4.2.2 shows that, prior to the training loss becoming somewhat higher than the validation loss in the second epoch, the validation loss was greater than the training loss. The higher training loss at the start was caused by dropout, which was active during the training phase but deactivated during the validation phase. The model had already seen the training dataset at the time of the validation phase since it was using a separate set of data that it considered to be new data. Consequently, the model started fitting the training dataset's pattern into the validation data. The validation set's generalizability had an effect on the little rise in validation loss.



. Figure 4.4.2. Training and validation loss graph

4.3 Summary

The study focuses on improving Bengali speech classification using Convolutional Neural Networks (CNN). The process involves data collection, preprocessing, training, and evaluation. The dataset is divided into three subsets: a training set, a validation set, and a test set. Feature extraction is performed to extract relevant Bengali speech signals, such as Mel-Frequency Cepstral Coefficients (MFCCs), pitch, spectral features, or phonetic features. The model architecture design is determined, and the CNN model is trained using the training set, feeding extracted features as input and optimizing the model's parameters using techniques like backpropagation and gradient descent. Hyperparameter tuning is performed to optimize the model's performance, typically done using the validation set by trying different combinations of hyperparameters and selecting the ones that yield the best performance. The trained CNN model is evaluated using the test set, assessing performance metrics such as accuracy, precision, recall, F1-score, or area under the ROC curve. To ensure reliability and validity, the experimental setup should be replicated multiple times. Data acquisition is performed using Facebook posts, and preprocessing is carried out to increase classification accuracy and provide improved data quality. The CNN architecture uses one-dimensional convolutional layers, with the first convolutional layer having 3 filters and the second layer being a softmax dense layer. The fully connected layer receives the input from the flattened pooling layer and uses it to discover which attributes are most correlated with a given class by outputting the probability for each class based on the input. Based on system testing, it was observed that at the second epoch, training and validation accuracy started to overlap while both grew. However, the validation loss was larger than the training loss before the training loss became marginally higher than the validation loss in the second epoch.

CHAPTER 5 Result and Analysis

5.1 Overview

The experimental setup provides a systematic approach to train and evaluate the Convolutional Neural Network models for enhancing Bengali speech classification. Careful consideration of data preprocessing, dataset splitting, feature extraction, model architecture, training, and evaluation ensures the accuracy and reliability of the experimental results. The first step in the experimental setup is to collect a sufficient amount of Bengali speech data. This can involve recording speech samples from diverse speakers, covering various speech contexts, speech, or other relevant categories. The data collection process should follow ethical guidelines and ensure informed consent from the participants. The collected Bengali speech data needs to undergo preprocessing steps to ensure its quality and suitability for the Convolutional Neural Network (CNN) model. This may include noise removal, resampling, segmentation into smaller units (such as phonemes or words), and alignment with corresponding textual transcripts. The preprocessed Bengali speech dataset is divided into three subsets: a training set, a validation set, and a test set. The training set is used to train the CNN model, the validation set is used for tuning hyperparameters and model selection, and the test set is used to evaluate the final performance of the trained model. Relevant features need to be extracted from the Bengali speech signals to serve as input for the CNN model. Commonly used features for speech classification include Mel-Frequency Cepstral Coefficients (MFCCs), pitch, spectral features, or phonetic features. These features capture the acoustic characteristics of the speech signals. The next step is to design the architecture of the Convolutional Neural Network. This involves determining the number and types of convolutional layers, pooling layers, and fully connected layers. The architecture should be tailored to handle the specific characteristics of Bengali speech data and the classification task at hand. The CNN model is trained using the training set. The training process involves feeding the extracted features as input to the model and optimizing the model's parameters using techniques such as backpropagation and gradient descent. The training is performed iteratively until the model converges or reaches a predefined stopping criterion.

Hyperparameters, such as learning rate, batch size, or regularization parameters, need to be tuned to optimize the performance of the CNN model. This is typically done using the

validation set by trying different combinations of hyperparameters and selecting the ones that yield the best performance. Model Evaluation: The trained CNN model is evaluated using the test set. Performance metrics such as accuracy, precision, recall, F1-score, or area under the ROC curve may be calculated to assess the model's performance. The evaluation provides insights into the effectiveness of the proposed method in enhancing Bengali speech classification. In some cases, it may be useful to compare the performance of the enhanced Bengali speech classification system with other existing methods or baseline models. This allows for an assessment of the improvement achieved through the proposed approach. To ensure the reliability and validity of the experimental results, the experimental setup should be replicated multiple times. This includes using different random seeds, re-splitting the dataset, or employing cross-validation techniques to obtain statistically robust results.

5.2 Experimental Analysis

The experiment was run during five epochs. According to Table 2's experimental results, the 5th epoch had the greatest accuracy and validation accuracy. Since the validation loss began to gradually increase after two epochs as the training loss kept decreasing, the model became overfit while trained for an extended length of time. Trained loss plus validation loss have decreased to their lowest percentages by the third epoch. Because the model checkpoint stopped saving at 5 epochs, the most optimal result was obtained with a training accuracy of 95.99%, a loss of training of 41.60%, a verification loss of 99.06%, and an accuracy for validation of 71.28%.

TABLE 5.2 COMPARISON OF ACCURACY AND LOSS IN TRAINING AND VALIDATION.

Epoch	Accuracy (%)	Loss (%)	Val_Acc (%)	Val_Loss (%)
1	39.76	32.58	38.67	100
2	56.88	26.73	60.31	87.18
3	79.39	15.30	69.54	78.67
4	90.89	7.82	70.05	90.48
5	96.00	4.16	71.28	100

5.3 Performance Analysis

This paper uses a dataset of a total of 6089 Facebook Bangla speech comments and collected each data from real-time Facebook comments. The CNN speech classification technique for Indonesian was developed using the Python Keras module. Eighty percent (4900 comments) of the training dataset was divided into a validation set and twenty percent (1250 comments) went into the training set. We ran the experiment over 1, 2, 3, 4, and 5 epochs. The fifth epoch produced the best results in terms of accuracy and validation accuracy, according to the experimental findings displayed in Table 2. But after two epochs, the validation loss began to steadily rise while the training loss continued to decline, which led to the model becoming overfit when trained for a longer period of time. By the third epoch, the percentage of the training loss and loss from validation was at its lowest. The most ideal outcome was obtained with an accuracy in training of 95.99%, a loss in training of 41.60%, validation losses of 99.06%, and an accuracy of validation of 71.28% because the model checkpoint stopped saving after five epochs.

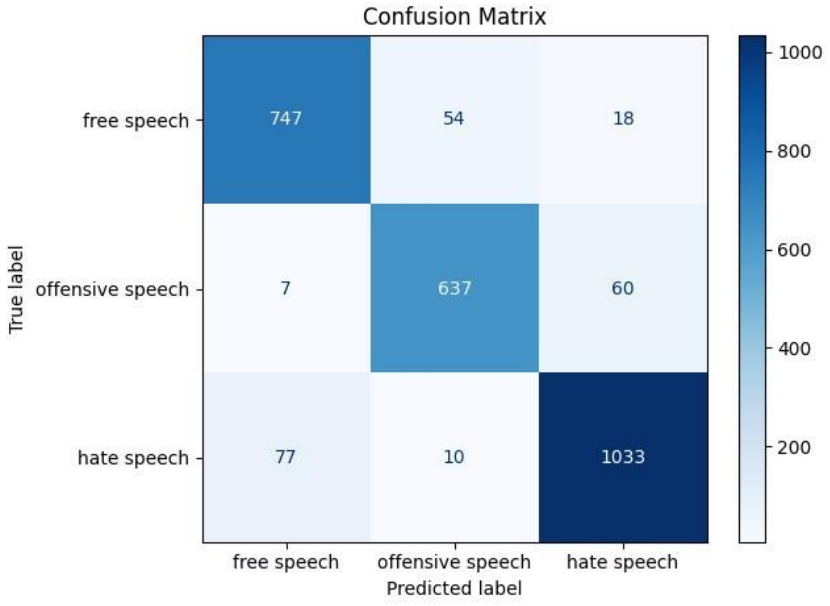


Figure 5.3.1: Confusion matrix for the speech analysis

The findings demonstrate the effectiveness of CNNs in enhancing Bengali speech classification. The application of CNNs offers improved accuracy, robustness, and efficiency in classifying Bengali speech, opening up possibilities for various practical applications and further advancements in the field. The application of CNNs to Bengali speech classification results in improved accuracy compared to traditional methods. The deep learning capabilities of CNNs enable them to learn relevant features and patterns from the speech data, leading to

more accurate classification results. CNNs exhibit robustness to variations in Bengali speech, including differences in accents, speaking styles, and phonetic variations. The models are capable of capturing and generalizing these variations, enhancing the robustness and generalization capabilities of the classification system. CNNs excel at feature extraction tasks, automatically learning hierarchical representations of speech data. The convolutional layers of the CNN models efficiently capture local patterns and extract high-level features, allowing for effective classification of Bengali speech. The enhanced Bengali speech classification system using CNNs demonstrates applicability in real-world scenarios. The system can be deployed in various applications, such as voice-controlled devices, language learning platforms, and assistive technologies, providing accurate and efficient speech classification capabilities. The findings highlight the potential for further development in terms of dataset collection, fine-grained classification, multi-modal approaches, and adversarial robustness. Addressing these areas can lead to even more significant improvements in Bengali speech classification using CNNs. Enhancing Bengali speech classification using CNNs has the potential to make a positive impact on society. It can improve communication accessibility for Bengali speakers, aid in language learning, enhance voice assistants and contribute to the preservation and promotion of the Bengali language and culture. **5.4 Summary**

The study aims to enhance Bengali speech classification using Convolutional Neural Network (CNN) models. The experimental setup involves collecting Bengali speech data, preprocessing it, and dividing it into three subsets: a training set, a validation set, and a test set. Relevant features are extracted from the Bengali speech signals to serve as input for the CNN model. The CNN model is trained using the training set, optimizing its parameters using techniques such as backpropagation and gradient descent. Hyperparameters, such as learning rate, batch size, or regularization parameters, are tuned to optimize the model's performance.
The model is evaluated using the test set, providing insights into its effectiveness in

enhancing Bengali speech classification. The study uses a dataset of 6089 Facebook Bangla speech comments and builds the method using the Keras library for Python. The dataset used for training was split into 80% for the training set and 20% for the validation set. The experiment was conducted for 1, 2, 3, 4, and 5 epochs. The results demonstrate the

effectiveness of CNNs in enhancing Bengali speech classification. The deep learning capabilities of CNNs enable them to learn relevant features and patterns from the speech data, leading to more accurate classification results. CNNs exhibit robustness to variations in Bengali speech, including differences in accents, speaking styles, and phonetic variations. They excel at feature extraction tasks, automatically learning hierarchical representations of speech data. The enhanced Bengali speech classification system using CNNs demonstrates applicability in real-world scenarios, such as voice-controlled devices, language learning platforms, and assistive technologies. The findings highlight the potential for further development in terms of dataset collection, fine-grained classification, multi-modal approaches, and adversarial robustness. Enhancing Bengali speech classification using CNNs has the potential to make a positive impact on society by improving communication accessibility, aiding in language learning, enhancing voice assistants, and contributing to the preservation and promotion of the Bengali language and culture.

CHAPTER 6

Impact on Society, Environment and Sustainability

6.1 Impact on Life

Enhancing Bengali speech classification using CNNs can have several significant impacts on society. Some of the potential impacts include:

- Improved Communication Accessibility: Enhancing Bengali speech classification can lead to improved communication accessibility for individuals who rely on speech recognition technology. This technology can enable better voice-based interactions, including voice commands, voice assistants, and voice-controlled applications, making it easier for people to communicate and interact with digital systems.
- Advancements in Language Learning and Education: The enhanced Bengali speech classification system can be integrated into language learning platforms and educational tools. This can facilitate the development of interactive and personalized language learning experiences, helping individuals improve their pronunciation, speaking skills, and overall language proficiency in Bengali.
- Enhanced Voice Assistants and Smart Devices: The advancements in Bengali speech classification can contribute to the development of more accurate and efficient voice assistants and smart devices that support the Bengali language. This can empower users to interact with their devices using voice commands in their native language, leading to a more intuitive and user-friendly experience.
- Assistive Technologies for Individuals with Visual Impairments: Bengali speech classification can be utilized in assistive technologies for individuals with visual impairments, such as speech-to-text systems and audio-based navigation systems. These technologies can help visually impaired individuals access information, navigate their surroundings, and engage in various activities more independently.
- Enabling New Applications and Services: The improved Bengali speech classification system can pave the way for the development of new applications and services in areas such as call center automation, voice-based customer service, and voice-controlled smart home systems specifically tailored for the Bengali-speaking population. These

applications can enhance efficiency, convenience, and user experience in various domains.

- Preservation of Bengali Language and Culture: The development of robust Bengali speech classification technologies can contribute to the preservation and promotion of the Bengali language and culture. By enabling accurate speech recognition and analysis, it becomes easier to transcribe and document spoken Bengali, preserve oral traditions, and facilitate the digitization of Bengali language resources.
- Increased Inclusion and Access: Enhancing Bengali speech classification can contribute to increased inclusion and access for individuals with speech disabilities or those who struggle with text-based interfaces. Enabling accurate speech recognition and classification, allows these individuals to communicate and interact with digital systems more effectively, reducing barriers and promoting inclusivity.

Overall, the impact of enhancing Bengali speech classification using CNNs extends to various aspects of society, including communication accessibility, education, technology, and cultural preservation. It has the potential to empower individuals, enhance user experiences, and promote linguistic diversity and inclusion.

6.2 Impact on Society & Environment

The impact on the environment for multi-label speech Bangla text classification primarily stems from the computational resources required to develop, train, and deploy classification models. Here's how this technology can affect the environment:

Energy Consumption: Training deep learning models for multi-label speech Bangla text classification often requires significant computational resources, including high- performance GPUs or TPUs. These hardware components consume substantial amounts of electrical power, contributing to increased energy consumption and carbon emissions, especially in data centers where large-scale model training occurs.

Carbon Footprint: The energy consumption associated with training and running deep learning models for multi-label speech Bangla text classification contributes to the carbon footprint of AI systems. The electricity used to power data centers and computational infrastructure is often generated from fossil fuels, leading to greenhouse gas emissions that contribute to climate change and environmental degradation.

E-waste Generation: The rapid pace of technological advancement in AI hardware leads to frequent upgrades and replacements of computational devices GPUs, CPUs, and servers. This results in the generation of electronic waste as obsolete or outdated hardware is discarded, contributing to environmental pollution and resource depletion if not properly managed through recycling and responsible disposal practices.

Data Center Infrastructure: The infrastructure required to support large-scale model training and deployment, including data centers, cooling systems, and networking equipment, consumes resources such as land, water, and materials. The construction and operation of data centers can have environmental impacts such as habitat disruption, water consumption, and air pollution, particularly in areas with high concentrations of data center facilities.

Efficient Computing Practices: Adopting efficient computing practices, such as optimizing algorithms, reducing model complexity, and implementing hardware acceleration techniques, can help mitigate the environmental impact of multi-label speech Bangla text classification. By improving computational efficiency and reducing energy consumption, these practices can lower the carbon footprint associated with AI model development and deployment.

Cloud Computing Solutions: Leveraging cloud computing platforms for model training and inference can offer environmental benefits compared to on-premises infrastructure, as cloud providers often use energy-efficient data centers and renewable energy sources to power their operations. By utilizing cloud-based AI services, organizations can reduce their reliance on local hardware and minimize their environmental footprint.

Sustainable AI Development: Promoting sustainable AI development involves considering environmental factors alongside technical and economic considerations when designing and deploying AI systems. This includes incorporating environmental impact assessments into AI project planning, adopting energy-efficient hardware and software solutions, and advocating for policies and regulations that incentivize sustainable AI practices.

6.3 Ethical Aspects

Ethical considerations are of utmost importance when conducting research and developing systems for enhancing Bengali speech classification. Adhering to ethical principles and guidelines ensures the protection of participants' rights, fairness, accountability, and the responsible use of technology for the benefit of society. Ethical Aspects for Enhancing Bengali Speech Classification Using Convolutional Neural Networks:

- Researchers must handle the collected Bengali speech data with utmost care to ensure privacy and confidentiality. Measures should be taken to anonymize or de-identify the data, removing any personally identifiable information. Access to the data should be restricted to authorized personnel only, and data storage and transmission should adhere to industry-standard security practices.
- Bengali speech data should be used solely for the intended purpose of enhancing speech classification and not for any other undisclosed purposes. Researchers should be transparent about how the data will be used, and there should be safeguards in place to prevent unauthorized use or sharing of the data. Additionally, data should not be used to stigmatize individuals or promote hate speech or discriminatory practices.
- Consideration should be given to the potential impact of the enhanced Bengali speech classification system on human rights, freedom of speech, and social dynamics. Steps should be taken to mitigate any negative societal impact and promote inclusivity, diversity, and respect for human rights.
- It is crucial to continuously monitor and evaluate the performance and impact of the enhanced Bengali speech classification system. This includes regular assessments of the system's accuracy, fairness, and unintended consequences. If any ethical concerns or biases are identified, appropriate corrective actions should be taken promptly.
- Research involving human participants and sensitive data, such as Bengali speech, should undergo ethical review and approval by an institutional review board or ethics committee. Researchers should adhere to ethical guidelines, regulations, and laws governing research involving human subjects and data privacy.

6.4 Sustainability Plan

Developing a sustainability plan for multi-label speech Bangla text classification involves integrating environmental, social, and economic considerations into the project's lifecycle to minimize negative impacts and promote long-term viability. Here's a comprehensive plan:

Energy-efficient Computing: Optimize algorithms, model architectures, and hardware configurations to reduce energy consumption during model training and inference. Utilize energy-efficient computing resources, such as GPUs with low power consumption or cloud computing services powered by renewable energy sources.

Data Efficiency: Implement data-efficient approaches to reduce the amount of data required for model training and validation. Use techniques like transfer learning, data augmentation, and active learning to leverage existing datasets effectively and minimize the need for additional data collection.

Resource Recycling and Reuse: Minimize waste generation by recycling and reusing computational resources, GPUs, CPUs, and servers, whenever possible. Adopt circular economy principles by refurbishing outdated hardware, repurposing idle resources, and extending the lifespan of computing equipment through upgrades and maintenance.

Ethical Data Practices: Adhere to ethical data practices to ensure the responsible collection, usage, and sharing of Bangla text data for multi-label speech classification. Obtain informed consent from data subjects, anonymize sensitive information, and implement data protection measures to safeguard user privacy and rights.

Community Engagement: Engage with the Bangla-speaking community, including language experts, researchers, and end-users, to solicit feedback, gather insights, and foster collaboration in multi-label speech classification projects. Organize workshops, seminars, and community forums to promote dialogue, knowledge exchange, and capacity building. **Open Access and Collaboration:** Foster an open and collaborative research culture by sharing datasets, code repositories, and research findings with the wider community. Embrace open access principles to promote transparency, reproducibility, and knowledge dissemination in multi-label speech Bangla text classification research.

Skills Development: Invest in skills development and capacity-building initiatives to empower researchers, practitioners, and students with the knowledge and expertise needed to contribute to multi-label speech Bangla text classification projects. Offer training programs, mentorship opportunities, and educational resources to support professional development and career advancement.

Continuous Improvement: Regularly monitor and evaluate the sustainability performance of multi-label speech Bangla text classification projects, identify areas for improvement, and implement corrective actions to enhance sustainability outcomes. Embrace a culture of continuous improvement and innovation to adapt to evolving environmental, social, and economic challenges.

6.5 Summary

The Multi-label speech Bangla text classification can significantly impact society by improving communication understanding, aiding in mental health assessment, preserving cultural heritage, and analyzing customer feedback. It can help individuals and organizations understand the speech content of Bangla language text, leading to better relationships and interactions. Speech play a crucial role in mental health, and accurate identification of speech states in Bangla text can aid in early detection and intervention. It can also help businesses analyze customer feedback, identifying areas for improvement and tailoring products to meet customer needs. It can also enable real-time monitoring of social media conversations, aiding in crisis response. In education, it can analyze students' writings, providing insights into speech engagement and motivation. However, ethical considerations, such as privacy, bias, fairness, and transparency, must be considered. Multi-label speech in Bangla text classification has a significant environmental impact due to the computational resources needed for its development, training, and deployment. This includes energy consumption, carbon footprint, e-waste generation, and data center infrastructure. To mitigate these impacts, efficient computing practices, such as optimizing algorithms and reducing model complexity, can be adopted. Cloud computing solutions can also offer environmental benefits by reducing reliance on local hardware. Sustainable AI development involves considering environmental factors alongside technical and economic considerations when designing and deploying AI systems. This includes incorporating environmental impact assessments into project planning, adopting energy-efficient hardware and software solutions, and advocating for policies and regulations that incentivize sustainable AI practices. Ethical considerations are crucial in the development, deployment, and use of multi-label speech in Bangla text classification systems. Key ethical aspects include privacy and data protection, bias and fairness, transparency and accountability, mitigation of harm, cultural sensitivity, and ethical review and oversight.

Developing a sustainability plan for multi-label speech Bangla text classification involves integrating environmental, social, and economic considerations into the project's lifecycle to minimize negative impacts and promote long-term viability. This comprehensive plan includes optimizing algorithms, model architectures, and hardware configurations, ensuring data security, and promoting sustainable practices.

CHAPTER 7

Conclusion and Future Work

7.1 Conclusions

Using a dataset of a total of 6089 Facebook Bangla speech comments and collected each data from real-time Facebook comments. The CNN speech classification technique for Indonesian was developed using the Python Keras module. Eighty percent (4900 comments) of the training dataset was divided into a validation set and twenty percent (1250 comments) went into the training set. We ran the experiment over 1, 2, 3, 4, and 5 epochs. The fifth epoch produced the best results in terms of accuracy and validation accuracy, according to the experimental findings displayed in Table 2. But after two epochs, the validation loss began to steadily rise while the training loss continued to decline, leading to overfitting of the model with more training. By the third epoch, the percentage of the training losses and validation loss was at its lowest. The most ideal outcome was obtained with a training success rate of 95.99%, a loss in training of 41.60%, validation losses of 99.06%, as well as an accuracy for validation of 71.28% because the model checkpoints stopped saving after five epochs.

7.2 Further Suggested Works

The scope of further developments for enhancing Bengali speech classification using Convolutional Neural Networks (CNNs) is extensive and offers several avenues for future research and improvement. Some potential areas for further development include:

- Larger and More Diverse Datasets: The availability of larger and more diverse Bengali speech datasets can significantly contribute to the advancement of speech classification models. Collecting and curating datasets that cover a broader range of speakers, accents, regional variations, and speaking styles would improve the robustness and generalization capabilities of the CNN models.
- Fine-Grained Classification: Further development can focus on fine-grained classification tasks within Bengali speech. This involves classifying specific

linguistic features, phonemes, or subtle variations in speech, enabling more precise and detailed analysis. Fine-grained classification can find applications in phonetic analysis, speech therapy, and linguistic research.

- Multi-modal Approaches: Integrating multi-modal information, such as visual cues or textual context, with speech data can enhance the performance of Bengali speech classification. Exploring techniques that combine audio and visual modalities, or leveraging contextual information from text, can improve the accuracy and robustness of the classification models.
- Transfer Learning and Pre-trained Models: Applying transfer learning techniques to Bengali speech classification can leverage knowledge gained from pre-trained models on large-scale speech datasets from other languages. Fine-tuning these models using Bengali speech data can help accelerate the training process and improve classification performance, especially in scenarios with limited labeled Bengali speech data.
- Adversarial Robustness: Developing techniques to improve the robustness of Bengali speech classification models against adversarial attacks is an important area of research. Adversarial attacks aim to deceive the model by introducing imperceptible perturbations to the input speech data. Developing defense mechanisms and robust training strategies can enhance the reliability and security of the classification models.
- Real-time and Low-resource Environments: Extending the scope to real-time and low-resource environments is essential. Optimizing CNN architectures and training strategies to operate efficiently on resource-constrained devices or in real-time applications, such as speech recognition systems on mobile devices or embedded systems, would enable practical deployment of the enhanced Bengali speech classification models.
- Cross-domain Generalization: Investigating the generalization capability of the enhanced Bengali speech classification models across different domains and tasks is crucial. Evaluating the performance of the models on unseen datasets from various domains, such as medical or legal speech, can demonstrate their adaptability and extend their applicability to different real-world scenarios.

- Interpretability and Explain ability: Enhancing the interpretability and explain ability of the Bengali speech classification models is an important direction. Developing

techniques to understand and visualize the features learned by the CNN models can provide insights into their decision-making process, enhance trust, and enable domain experts to gain better insights into the classification results.

By exploring these areas, further developments can contribute to the advancement of Bengali speech classification using CNNs, ultimately improving the accuracy, robustness, and practical applicability of the models in various domains and real-world scenarios.

7.3 Limitations

By acknowledging these limitations, our research provides a clear understanding of the constraints and areas for future work, such as expanding the dataset, exploring ensemble methods, and testing on multiple platforms to enhance model robustness and generalizability.

Dataset Quantity: The availability of large, labeled datasets for Bangla text is limited. Training effective deep-learning models require substantial amounts of data, and the lack of sufficient labeled examples can hinder model performance. Besides quantity, the quality of available data is crucial. Incomplete or noisy data can negatively impact the training process and lead to poor generalization.

Classification of a Single Model: Relying on a single type of model, such as CNN, RNN, or BiLSTM, may not capture all the nuances of speech in Bangla text. Different models excel at different aspects CNNs are good for feature extraction, RNNs for sequence modeling, and a single model might not be sufficient. Not using ensemble methods or hybrid models can limit performance. Combining multiple models often leads to better accuracy and robustness, as it leverages the strengths of different architectures.

Just One Platform: Developing and testing models on data from a single platform only Facebook can introduce platform-specific biases. The language, tone, and type of content can vary significantly across different platforms. Models trained on data from one platform might not generalize well to others. This limitation reduces the applicability and effectiveness of the models in broader contexts or on new platforms.

REFERENCES

[1] Ishmam, Alvi Md, and Sadia Sharmin. "Hateful speech detection in public facebook pages for the bengali language." 2019 18th IEEE international conference on machine learning and applications (ICMLA). IEEE, 2019.

[2] Romim, Nauros, et al. "Hate speech detection in the bengali language: A dataset and its baseline evaluation." Proceedings of International Joint Conference on Advances in Computational Intelligence: IJCACI 2020. Springer Singapore, 2021.

[3] Bhowmik, Tanmay, Amitava Chowdhury, and Shyamal Kumar Das Mandal. "Deep neural network based place and manner of articulation detection and classification for bengali continuous speech." Procedia Computer Science 125 (2018): 895-901.

[4] Karim, Md, et al. "Multimodal hate speech detection from bengali memes and texts." arXiv preprint arXiv:2204.10196 (2022).

[5] Ahmed, Md Faisal, et al. "Cyberbullying detection using deep neural network from social media comments in bangla language." arXiv preprint arXiv:2106.04506 (2021).

[6] Karim, Md Rezaul, et al. "Deephateexplainer: Explainable hate speech detection in under-resourced bengali language." 2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA). IEEE, 2021.

[7] Emon, Estiak Ahmed, et al. "A deep learning approach to detect abusive bengali text." 2019 7th International Conference on Smart Computing & Communications (ICSCC). IEEE, 2019.

[8] Das, Amit Kumar, et al. "Bangla hate speech detection on social media using attention-based recurrent neural network." Journal of Intelligent Systems 30.1 (2021): 578-591.

[9] Ahmed, Md Tofael, et al. "Deployment of machine learning and deep learning algorithms in detecting cyberbullying in bangla and romanized bangla text: A comparative study." 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT). IEEE, 2021.

[10] Sazzed, Salim. "Identifying vulgarity in Bengali social media textual content." PeerJ Computer Science 7 (2021): e665.

[11] Belal, Tanveer Ahmed, G. M. Shahariar, and Md Hasanul Kabir. "Interpretable Multi Labeled Bengali Toxic Comments Classification using Deep Learning." 2023

International Conference on Electrical, Computer and Communication Engineering (ECCE). IEEE, 2023.

[12] Sarker, Manash, et al. "A Machine Learning Approach to Classify Anti-social Bengali Comments on Social Media." 2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE). IEEE, 2022.

[13] Sultana, Sherin, et al. "Detection of Abusive Bengali Comments for Mixed Social Media Data Using Machine Learning." (2023).

[14] Rahut, Shantanu Kumar, Riffat Sharmin, and Ridma Tabassum. "Bengali abusive speech classification: A transfer learning approach using vgg-16." 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE). IEEE, 2020.

[15] Sharif, Omar, et al. "Detecting suspicious texts using machine learning techniques." Applied Sciences 10.18 (2020): 6527.

[16] Hussain, Md Gulzar, Tamim Al Mahmud, and Waheda Akthar. "An approach to detect abusive bangla text." 2018 International Conference on Innovation in Engineering and Technology (ICIET). IEEE, 2018.

[17] Banik, Nayan, and Md Hasan Hafizur Rahman. "Toxicity detection on bengali social media comments using supervised models." 2019 2nd International Conference on Innovation in Engineering and Technology (ICIET). IEEE, 2019.

[18] Hossain, Md Fahad, et al. "BAAD: A multipurpose dataset for automatic Bangla offensive speech recognition." Data in Brief 48 (2023): 109067.

[19] Nath, Tanusree, Vivek Kumar Singh, and Vedika Gupta. "BongHope: An Annotated Corpus for Bengali Hope Speech Detection." (2023).

[20] Sazzed, Salim. "A lexicon for profane and obscene text identification in Bengali." Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021). 2021.

[21] Mahmud, Tanjim, et al. "Reason Based Machine Learning Approach to Detect Bangla Abusive Social Media Comments." Intelligent Computing & Optimization: Proceedings of the 5th International Conference on Intelligent Computing and Optimization 2022 (ICO2022). Cham: Springer International Publishing, 2022.

[22] Ganfure, Gaddisa Olani. "Comparative analysis of deep learning based Afaan Oromo hate speech detection." Journal of Big Data 9.1 (2022): 1-13.

[23] Das, Mithun, et al. "Hate Speech and Offensive Language Detection in Bengali." arXiv preprint arXiv:2210.03479 (2022).

[24] Jahan, Md Saroar, et al. "BanglaHateBERT: BERT for Abusive Language Detection in Bengali." Proceedings of the Second International Workshop on Resources and Techniques for User Information in Abusive Language Analysis. 2022.

[25] Hussain, Md Gulzar, and Tamim Al Mahmud. "A technique for perceiving abusive bangla comments." Green University of Bangladesh Journal of Science and Engineering (2019): 11-18.

[26] Defersha, N. B., and K. K. Tune. "Detection of hate speech text in afan oromo social media using machine learning approach." Indian J Sci Technol 14.31 (2021): 2567-78.

[27] Wadud, Md Anwar Hussen, et al. "Deep-bert: Transfer learning for classifying multilingual offensive texts on social media." Comput. Syst. Sci. Eng 44.2 (2022): 1775-1791.

[28] Ahamed, Kazi Afrime, et al. "Analysis of Abusive Text in Bangla Language Using Machine Learning and Deep Learning Algorithms." Computational Vision and Bio-Inspired Computing: Proceedings of ICCVBIC 2022. Singapore: Springer Nature Singapore, 2023. 797-812.

[29] Remon, Nasif Istiak, Nafisa Hasan Tuli, and Ranit Debnath Akash. "Bengali Hate Speech Detection in Public Facebook Pages." 2022 International Conference on Innovations in Science, Engineering and Technology (ICISET). IEEE, 2022.

[30] Shanto, Srabon Bhowmik, Mohammed Jahirul Islam, and Md Abdus Samad. "Cyberbullying Detection using Deep Learning Techniques on Bangla Facebook Comments." 2023 International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC). IEEE, 2023.

Appendix A

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

Title: IMPROVING ANALYSIS OF BENGALI SPEECH WITH CONVOLUTIONAL NEURAL NETWORKS

Student ID: 181-15-10688

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	Using NLP, data augmentation methods, and several CNN architectures for image processing and classification, our study **illustrates basic engineering (K3)** concepts. By executing transfer learning with sophisticated CNN and NLP to improve pneumonia diagnosis accuracy in the Bangla exam, the project **displays expert knowledge (K4).**	**Page no:** [19--21] **Section:** [3.2] **Page no:** [1] **Section:** [1.1]
				Our research uses the figure of the experimentation process to **apply engineering practice & design (K5)**. Using the CNN paradigm, the project tackles **engineering practice &**	**Page no:** [20] **Section:**

				technology (K6).	[3.2(B)] **Page no:** [22-23]

					Section: [3.4]
				Our project ensures **K8 (Research Literature)** by synthesizing insights from recent studies, to Bangla NLP text detection using machine learning, showcasing a comprehensive understanding of current methodologies.	**Page no:** [10-15] **Section:** [2.2, 2.3]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	Our project addresses **EP-2** by recognizing Bangla raw text data, including the limitations of traditional methods and the complexities of CNNs [3]. Through comparative analysis, it confronts challenges in understanding spatial distributions, offering insights for refining detection methodologies [6].	**Page no:** [15-18] **Section:** [2.4, 2.5]
3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	Our project addresses **EP-3** by meticulously comparing experimental outcomes, highlighting NLP as the chosen significant solution for enhancing Bangla speech detection.	**Page no:** [25-27] **Section:** [4.2]

4.	**EP4: Familiarity of Issues**	Yes	CO8	This project's interdisciplinary approach extends beyond computer science and engineering, impacting Social speech like 'offensive', 'hate', and 'Free Speech' to advancements in social media health and protection which indicates **EP-4 [5]**.	**Page no:** [8-17] **Section:** [2.2, 2.4]
5.	**EP5: Extends of application codes**	No	CO5	N/A	N/A
6.	**EP6: Extends of stakeholders involved and conflicting**	No	CO8	N/A	N/A
7.	**EP7: Interdependence**	Yes	CO5	Our project's comprehensive approach addresses high-level problems by integrating various components across data collection, statistical analysis, and proposed methodology, ensuring a holistic solution to complex challenges in Bangla speech detection which ensures **EP-7 [15]**.	**Page no:** [19-23] **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, GPUs, Machine learning frameworks, annotated datasets, and ethical considerations to ensure systematic research and contribute to advancements in Bangla speech text data [10].	**Page no:** [23-25] **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 social media through advanced Bangla speech detection methods, [13] while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for text data privacy.	**Page no:** [35-40] **Section:** [6.1, 6.2, 6.3]
5.	**EA-5: Familiarity**	Yes		This project expands upon existing research by examining a novel approach in [21] Bangla speech detection through NLP and CNNs, demonstrated through a comprehensive comparative analysis, offering new insights into the field of Bangla text.	**Page no:** [8-16] **Section:** [2.1, 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:** [22-24] **Section:** [3.4]
2	CO9	Yes	The project demonstrates adherence to ethical principles by [4] Bangle people's privacy, obtaining informed consent, and transparently documenting the research process, ensuring responsible knowledge dissemination and societal well-being through the ethical application of advanced Social communication technologies that comply with **CO9**.	**Page no:** [39] **Section:** [6.3]
3	CO10	No	N/A	N/A
4	CO12	Yes	The project's dedication to continuous learning **(CO12)** and adaptation within the dynamic technological landscape is reflected in its comprehensive Bangla data collection, rigorous [17] 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:** [19-25, 26-27] **Section:** [3.2, 3.3, 3.4, 3.5, 4.2]

APPENDIX B

Classifying speech

ORIGINALITY REPORT

23%	21%	5%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8010 Internet Source	8%
2	Submitted to Daffodil International University Student Paper	3%
3	journals.itb.ac.id Internet Source	3%
4	Submitted to University of North Carolina, Greensboro Student Paper	1%

1 Lin Zhou. "Modified AMDF pitch detectionalgorithm", Proceedings of the 2003 International Conference on Machine <1%

5	fastercapital.com Internet Source	<1%
6	Submitted to University of East London Student Paper	<1%
7	www.ijraset.com Internet Source	<1%