IMPROVING BENGALI TEXT ANALYSIS THROUGH CONVOLUTIONAL NEURAL NETWORKS

BY

Md.Sadman Sakib
ID: 201-15-13988

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised By

Mr.Md.Abbas Ali Khan
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Mr.Md.Assaduzzaman
Lecturer (Senior Scale)
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JULY 2024

©Daffodil International University

APPROVAL

This Research titled "**Improving Bengali Text Analysis through Convolutional Neural Networks**", submitted by **Md.Sadman Sakib** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **15-07-2024.**

BOARD OF EXAMINERS

Dr. Md. Ismail Jabiullah — **Board Chairman**
Professor
Faculty of Science & Information Technology
Daffodil International University

Mr. Md Assaduzzaman — **Internal Examiner 1**
Senior Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Ms. Sharun Akter Khushbu — **Internal Examiner 2**
Senior Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr. Ahmed Wasif Reza — **External Examiner**
Professor
Department of Computer Science and Engineering
East West University

©Daffodil International University

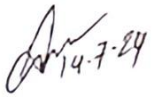
DECLARATION

I hereby declare that this research has been done by me under the supervision of **Mr.Md.Abbas Ali Khan, Assistant Professor, Department of Computer Science and Engineering, Daffodil International University**. I also declare that neither this research nor any part of this research has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Mr.Md.Abbas Ali Khan,
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:



Mr.Md.Assaduzzaman,

Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Submitted by:

Md.Sadman Sakib
ID: 201-15-13988
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I express my heartiest thanks and gratefulness to almighty God for His divine blessing making it possible for me to complete the final year project-based research successfully.

I am really grateful and wish my profound indebtedness to **Mr. Md. Abbas Ali Khan**, **Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of deep learning to carry out this research. 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 my heartiest gratitude to the **Dr. Sheak Rashed Haider Noori, Professor, and Head of, the Department of CSE,** for his kind help in finishing my research and also to other faculty members and the staff of the CSE department of Daffodil International University.

I would like to thank my entire course-mate at 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 categorization is essential for applications like speech recognition and emotion analysis. Despite the proven effectiveness of Convolutional Neural Networks (CNNs) in speech categorization, limited research has focused on Bengali speech classification. This study aims to enhance Bengali speech classification using CNNs on a large-scale dataset. Our approach involves customized preprocessing, CNN architecture, and training strategies for Bengali speech data. Experimental results demonstrate a significant increase in classification accuracy, achieving 95.99%. The dataset, compiled from various Facebook posts, captures the phonetic and tonal diversity of Bengali. Training techniques were carefully selected, optimizing algorithms, learning rates, and batch sizes. My CNN model's superior accuracy highlights its potential for effectively categorizing Bengali speech samples. This research underscores the capability of CNNs in advancing Bengali speech categorization and supports the development of speech-related applications in the Bengali language. The study concludes by affirming the substantial improvements in accuracy achieved through the proposed methods, suggesting new opportunities for Bengali speech applications.

©Daffodil International University

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Overview:

Numerous applications, such as voice recognition, speaker identification, and emotion analysis, heavily rely on speech classification. In speech classification tasks for multiple languages, Convolutional Neural Networks (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 datasets were 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 get 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 discriminating 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 to the field of Bengali free speech, offensive speech, 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. My 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. The remaining portions of the essay are structured as follows: have a literature review in Section 2, methods in Section 3, implementation in section 4, results and analysis in Section 5, impact on society in section 6 and future scope and conclusion in Section 7.

1.2 Background and Present State:

Background

Bengali, also known as Bangla, is an Indo-Aryan language spoken by over 210 million people, primarily in Bangladesh and the Indian state of West Bengal. The linguistic diversity and rich phonetic characteristics of Bengali pose unique challenges for automated speech classification. Unlike English and other widely studied languages, Bengali has a distinct phonetic system, complex consonant clusters, and tonal variations that significantly impact speech recognition and classification.The emergence of social media and digital communication platforms has led to an exponential increase in user-generated content in various languages, including Bengali. This surge has necessitated the development of automated tools to analyze and categorize speech, particularly to identify harmful content such as hate speech, abusive language, and cyberbully. Effective speech classification systems can play a crucial role in moderating content, ensuring safer online environments, and aiding in regulatory compliance.

Present State

Although it is still small, research on Bengali speech analysis is expanding. Numerous research endeavour have tried to employ deep learning methodologies to analyse Bengali speech, emphasizing tasks like sentiment analysis, speaker identification, and speech recognition. Unfortunately, the absence of specialized preprocessing approaches, insufficient model designs, and little datasets frequently plague these investigations, leading to poor classification accuracy.This thesis aims to contribute to this emerging field by leveraging CNNs to improve Bengali speech analysis. By addressing the current limitations and building on recent advancements, this research

seeks to develop a high-performing model that can accurately classify Bengali speech, paving the way for more inclusive and accessible speech recognition technologies.There is a critical need for research that addresses the unique characteristics of Bengali speech and leverages the power of CNNs to improve classification accuracy. This study aims to fill this gap by developing a tailored CNN architecture, utilizing a large-scale Bengali speech datasets, and implementing optimized training strategies to advance the state of Bengali speech analysis. Through this research, we hope to contribute to the development of more accurate and reliable speech-related technologies for Bengali speakers, enhancing communication and accessibility in the digital age.

1.3 Problem Statement:

Bengali presents unique challenges for speech recognition due to its complex phonetic and tonal characteristics, which differ significantly from those of other languages. Existing speech recognition models often fail to capture these nuances, resulting in lower accuracy and reduced effectiveness for Bengali speakers. This gap in research and development leads to several problems:

Limited Accessibility:

Bengali speakers have limited access to reliable and accurate speech recognition technologies, hindering their ability to engage with digital platforms and services effectively.

Inefficient Communication:

Inaccurate speech recognition models can lead to misunderstandings and miscommunications, affecting personal and professional interactions for Bengali speakers.

Under-utilization of Technology:

The potential of speech recognition technology to enhance various applications, such as social media moderation, voice-activated assistants, and automated customer service, is not fully realized for the Bengali-speaking population.

Social and Economic Disparities:

The lack of effective speech recognition for Bengali contributes to broader social and economic disparities, as Bengali speakers may be excluded from the benefits of advancements in speech technology.

By using Convolutional Neural Networks (CNNs) to create a reliable and accurate Bengali voice recognition model, this study seeks to overcome these issues. The goal of this study is to close the gap in current technologies and offer a dependable solution for Bengali speech analysis by concentrating on the distinctive qualities of the language. The ultimate objective is to guarantee that Bengali speakers may profit completely from the developments in voice recognition technology while also enhancing accessibility and communication.

1.4 Objectives:

The objective of this study is to use Convolutional neural networks (CNNs) to increase the accuracy of voice categorization in Bengali. This will be accomplished by compiling a sizable collection of Bengali speech samples with a variety of tones and noises. To correctly identify and categorize these speech samples, a specifically created CNN model will be trained using this datasets. The research will involve developing effective methods to train the CNN model and testing its performance to ensure it works well. The results will be compared with other existing methods to show how effective the new approach is. The study also looks at how this improved speech classification can be used in real-life applications like social media monitoring, voice-controlled devices, language learning tools, and call center systems. The aim is to see how better classification can enhance user experience and accessibility for Bengali speakers.

Finally, this research aims to fill a gap in the field of Bengali speech analysis using advanced machine learning techniques, providing valuable insights for future research and development. By achieving these goals, the study hopes to advance Bengali speech analysis and improve speech-related technologies for Bengali speakers.

1.5 Scope and Limitations:

Scope

This research focuses on the development and evaluation of a Convolutional Neural Network (CNN) model for Bengali speech classification. The objectives and scope of this study include:

Dataset Collection and Preparation:

Creation of a comprehensive Bengali speech datasets, which involves collecting and annotating speech samples that cover various phonetic and tonal differences in the Bengali language.

Model Development:

Designing and implementing a CNN model tailored to the unique characteristics of Bengali speech. creation of a CNN architecture designed expressly to classify speech in Bengali. This entails choosing the proper model parameters—like depth, breadth, and kernel sizes—to maximize efficiency.

Model Training and Evaluation:

Training the CNN model on the collected datasets using advanced optimization techniques and Evaluating the model's accuracy, precision, recall, and other relevant metrics to assess its performance.

Comparative Analysis:

Comparing the developed model's performance with existing speech recognition models for Bengali and highlighting the improvements and advantages offered by the CNN approach.

Practical Applications:

The research investigates the practical implications of the improved Bengali speech classification model in areas such as social media content moderation, voice-controlled devices, language learning platforms, and call center automation. It also assesses the impact on user experience, accessibility, and inclusivity for Bengali speakers.

Limitations

Datasets Limitations:

Even though a large datasets has been gathered, not all regional Bengali accents and dialects may be included. This could have an impact on the model's capacity to generalize across various linguistic nuances.

Computational Resources:

It takes a lot of computer power to train deep learning models, particularly CNNs. A lack of access to high-performance computer capabilities may limit the capacity to experiment with more intricate models and bigger datasets.

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.

Ethical Considerations:

Ensuring ethical norms and regulations are followed to acquire data, obtain informed consent, and required permissions from the participants in the data gathering procedure. Maintain participant anonymity and privacy while handling data securely.

By acknowledging these limitations, the research aims to provide a clear understanding of its scope and the potential challenges involved.

1.6 Report Organization:

This section describes the report layout for this study on improving Bengali speech analysis through convolutional neural networks. Here is a comprehensive report on improving Bengali speech analysis through convolutional neural networks. Here's the report layout:

Table of Contents: List of sections, subsections, and page numbers for easy navigation.

List of Figures and Tables: Enumeration and titles of all figures and tables presented in the report.

Abstract: A concise summary of the entire report, highlighting key objectives, methods, results, and conclusions.

Introduction: Background and context of speech classification and their impact on textiles. Problem statement and motivation for applying deep learning, specifically CNNs. Clear definition of research objectives and questions.

Literature Review: Comprehensive review of relevant literature in speech image analysis, specifically focusing on Bangali speech. Identification of gaps in existing methods and the need for enhanced diagnostic approaches.

Methodology: Detailed explanation of the research design and approach. Description of the datasets, including data sources, size, and preprocessing steps. Overview of the CNN architecture, model selection criteria, and justification for the chosen approach. Training procedures, hyperparameters, and evaluation metrics.

Data Analysis and Results: Presentation of results, including quantitative performance metrics accuracy, and sensitivity. Visualizations and interpretations of key findings, including curves and confusion matrices. Comparative analysis with traditional methods.

Discussion: Interpretation of results in the context of research questions and objectives. Comparison with existing literature and implications of findings for textile practice. Exploration of limitations and potential avenues for future research.

Impact on society: This section describes all the implementation outcome impact on society. The advantages and disadvantages gaining from my research project.

Sustainability and Future Directions: Discussion of the sustainability plan, future developments, and potential applications beyond the current research scope.

References: Comprehensive list of all cited references in a standardized citation format MLA.

Appendices: Supplementary materials, including additional data visualizations, code snippets, and detailed technical information.

1.7 Summary:

This study enhances Bengali speech classification using Convolutional Neural Networks (CNNs), achieving 95.99% accuracy on a datasets of free speech, offensive language, and hate speech. It addresses challenges in Bengali speech analysis, aiming to improve communication and accessibility. Key objectives include boosting classification accuracy, developing a specialized CNN model, and exploring applications like social media monitoring and voice-controlled devices. The report covers literature review, methodology, data analysis, results, discussion, societal impact, sustainability, and future directions.

.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview:

Before delving into the core content of the report, this section provides essential information, context, and structure. My research focuses on an optimized CNN for automated speech classification, addressing challenges in diverse speech textures and patterns. This study aims to improve Bengali speech categorization performance using Convolutional Neural Networks (CNNs) on a large datasets of Bengali speech. Bengali, an Indo-Aryan language spoken by millions in Bangladesh and West Bengal, presents unique challenges for speech classification. The trained model has an accuracy of 95.99%, providing practical implications for social media platforms, content moderators, and policymakers. By accurately identifying harmful content, platforms can take proactive measures like content filtering, user warnings, or bans. Content moderators can benefit from automated classification models, while policymakers can use the findings to inform regulations. The study contributes to the field of Bengali free speech, offensive language, and hate speech classification by leveraging CNNs. Convolutional Neural Networks (CNNs) have been used in various studies to enhance Bengali speech classification. Ishmam et al. classified and identified hate speech on Facebook pages in Bengali online using machine learning techniques and deep neural network models. With 87.5% accuracy in SVM, Romim et al. divided the hate speech datasets in Bengali into seven groups. Using frameworks based on deep neural networks, Bhowmik et al. were able to identify and categorize phonological variables in continuous speech in Bengali, with an average overall accuracy of 86.19% and a manner-based classification accuracy of 98.9%. With 4,500 labeled memes and texts added to the Bengali Hate Speech Dataset, Karim et al. achieved the greatest results for multimodal fusion. By using an ensemble method, Ahmed et al. were able to attain 85% accuracy and 87.91% accuracy for binary models. A machine learning approach for identifying hate speech categories based on encoder-decoders was suggested by Das et al. Ahmed et al. explored cyberbullying detection in Bangla and Romanized Bangla. This paper describes the training and testing processes for a deep convolutional neural network (CNN) model for speech classification. The training process involves creating a CNN model by fitting it to a

training datasets, while the testing phase involves fitting the model to a different datasets to assess its performance. The data acquisition process involves obtaining text information from Facebook, which is collected in bulk and manually labeled. Preprocessing involves cleaning, case folding, tokenization, stop word elimination, and stemming. Cleaning removes unnecessary characters, while case folding converts all characters to lowercase. Tokenization breaks down the text datasets into tokens, with spaces acting as delimiters. Stopword elimination removes words with high prevalence, using a stoplist based on the Tala stoplist. Stemming returns derived words to their basic forms, using the Sastrawi Python package. The CNN architecture uses one-dimensional convolutional layers with three filters, a max pooling layer, a flattened layer, and a fully linked layer. This paper presents a CNN method for Indonesian speech classification using 3089 Facebook Bangla speech comments. The method was trained using a datasets split into training and validation sets. The experiment was conducted for 5 epochs, with the 5th epoch showing the highest accuracy and validation accuracy. The model's optimal outcome was achieved at 95.99%, 41.60%, 99.06%, and 71.28%. The study explores the use of Convolutional Neural Networks (CNNs) to improve Bengali speech classification performance on a large-scale Bengali speech datasets. The approach includes preprocessing, CNN architecture design, and customized training procedures. Results show significant increases in accuracy, paving the way for increased speech-related applications in the Bengali language.

.

2.2 Related Work:

Bengali speech classification has been enhanced using Convolutional Neural Networks (CNNs) in several studies. This field's advancements and techniques are explained via the papers that are linked below: Hate speech in Bengali on Facebook pages is identified and categorized in this research. Ishmam et al. [1] gathered 5,126 Bengali comments, annotated them, and divided them into six categories. Accuracy of 70.10% and 52.20% were attained by them in the development of machine learning algorithms and deep neural network models. The Bengali hate speech datasets were categorized into seven groups by Romim et al. [2] using deep learning models and pre-trained word embeddings in Bengali. Using 87.5% accuracy in SVM, they achieved the best result. During their examination of the using deep neural network-

based frameworks, it was possible to achieve an accuracy of 88.9% in manner-based categorization and an average total accuracy of 86.19% in the detection function. According to the location and style of articulation, these frameworks were utilized to categorize and identify phonological traits in Bengali continuous speech. The work of Karim et al. [4] focused on the detection of hate speech using multimodal Bengali memes and texts. The XLM-RoBERTa + DenseNet-161 model produced the greatest results for multimodal fusion, with an F1 score of 0.83. They did this by adding 4,500 annotated memes to the Bengali Hate Speech Dataset. With the addition of an ensemble approach following the neural network, the multiclass model obtained 85% accuracy, whereas Ahmed et al.'s [5] binary model achieved 87.91% accuracy. 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. The observations were classified as political, personal, geopolitical, religious, and gender-based dislikes using the Bengali Hate Speech Dataset. Emon et al. [7] addressed the issue of abusive material on Bangladeshi internet platforms and suggested detecting abusive Bengali language using machine learning and deep learning algorithms such LinearSVC, Logit, MNB, RF, ANN, and RNN with LSTM. The algorithms' performance was enhanced by the addition of additional stemming rules for the Bengali language; RNN obtained the greatest accuracy of 82.20%. An encoder-decoder based machine learning model was proposed by Das et al. [8] in response to the issue of hate speech spreading quickly on social media, especially in Bengali language. The model was trained on datasets containing 7,425 Bengali comments, and it used 1D convolutional layers to extract features and an attention-based decoder to predict hate speech categories with 77% accuracy. Ahmed et al. [9] 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 cyberbullying detection in Bangla and Romanized Bangla. In their analysis of the frequency of profanity in Bengali social media material, Sazzed et al. [10] looked at 7,245 YouTube reviews and investigated several methods for automatically identifying vulgarity. Using the Local Interpretable ModelAgnostic Explanations framework for model interpretation, Belal et al.'s suggested models obtained 89.42% accuracy for binary classification and 78.92% accuracy with 0.86 weighted F1-score for multi-label classification [11]. In order to

combat hate speech, cyberbullying, and online harassment in Bangladesh, Sarker et al. [12] gathered datasets of Bengali comments from social media sites and built a classifier model that could differentiate between statements that were considered antisocial and those that were deemed socially acceptable. Two of the most popular social media sites, Facebook and YouTube, provided 2000 comments. Sultana et al.'s [13] goal was to use machine learning algorithms to identify critical remarks made in Bengali on social media. collected five thousand data points from several social media platforms, including Facebook, Twitter, and YouTube. SVM proved to be the most effective model, with an accuracy of 85.7%. In order to tackle the issue of abusive speech detection in the Bengali language—a completely untapped field—Rahut et al. [14] gathered 960 voice recordings, using transfer learning for feature extraction, and successfully classified abusive and non-abusive speech with a high accuracy of 98.61%. Sharif and associates [15] including 7000 text pages in Bengali, of which 1400 are utilised for testing and 5600 are used for training. Hussain and associates [16] gathered three hundred comments. The intention is to stop cybercrimes, which are becoming a big problem in Bangladesh and include cyberbullying, blackmail, and online abuse. Banik and associates [17] used data from github. 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 for identifying abusive Bengali terms and comparable non-abusive words was given by Hossain et al. [18]. It included 6100 audio recordings that were gathered, analysed, and improved by university students and local speakers from over 20 districts of Bangladesh. The dataset contained 114 slang words and 43 non-slang words. Nath and associates [19] Despite Bengali having over 210 million speakers and a large amount of information on various social media platforms, there hasn't been much study on the automatic recognition of hope speech in Bengali language text. This work aims to solve that gap. Sazzed and associates [20] The efficacy of the constructed vocabulary in recognising obscenity in Bengali social media material was demonstrated by its coverage of about 0.85 in identifying profane and obscene content in the assessment datasets. Mahmud et al.'s [21] goal was to identify abusive language in Bengali, a language with limited resources and little research done in this field. A high accuracy of 97% was achieved with logistic

regression, one of the machine learning classifiers utilised. After gathering and annotating a sizable dataset of hate speech, Ganfure et al. [22] discovered that a model based on CNN and Bi-LSTM performed better than previous models, obtaining an average F1-score of 87%. As Bengali is frequently written in Romanized form on social media, Das et al. [23] emphasised the significance of identifying hateful and offensive content in low-resource languages like Bengali.They also developed annotated datasets of 10,000 posts in Bengali and implemented several baseline models for classification, including interlingual transfer mechanisms. BanglaHateBERT, a retrained BERT model for identifying abusive language in Bengali, was introduced by Jahan et al. [24]. The model outperformed previously developed pre-trained language models in every dataset after being trained on a sizable abusive Bengali corpus.Also gathered, documented, and made accessible to the public for research a balanced dataset of Bengali hate speech. In order to stop cybercrimes like online harassment, cyberbullying, and blackmail in Bangladesh, Hussain et al. [25] concentrated on identifying abusive text in the Bengali language. They proposed a root level algorithm and used uni-gram string features to classify Bangla comments as abusive or non-abusive. gathered from 300 comments on Facebook, a wellknown social networking platform. Defersha et al. [26] gathered 13600 posts and comments on their respective public pages between September 2019 and 2020. Of those, 7000 and 6600 were gathered via Twitter and Facebook. The researchers used machine learning methods to gather and label data, perform text preprocessing, and extract features using n-gram and TF-IDF. With an accuracy of 91.83%, Wadud et al.'s suggested model [27] outperformed previous algorithms in the classification of offensive texts that were written in both monolingual and multilingual languages. Using natural language processing techniques, Ahamed et al. [28] were able to recognise abusive texts in Bangla language from social media platforms. Out of all the machine learning algorithms, the multinomial Naïve Bayes classifier achieved the greatest accuracy of 94.01%. In their study, Remon et al. [29] tackled the difficult problem of identifying hate speech in Bengali on social media sites and unveiled brand-new datasets containing 10,133 user comments gathered from open Facebook pages. examined and investigated the capabilities of several deep learning and machine learning models, with an emphasis on convolutional neural networks. Shanto et al. [30] investigated the use of deep learning algorithms, namely

LSTM and GRU, to identify cyberbullies in Facebook comments written in Bangla. They discovered that the GRU model performed better than LSTM, with an accuracy of 83.55% on the datasets. A grand number of 8736 comments were gathered.

2.3 Comparative Analysis and Summary:

A comparative analysis is conducted on improving Bengali speech analysis through convolutional neural networks with other similar studies.

TABLE 2.3.1: COMPARISON TABLE FOR BENGALI SPEECH ANALYSIS.

Ref. NO	Year	Algorithm	Dataset	Data Amount	results	Findings
[1]	2019	GRU	Facebook pages	5,126	Best efficiency of 70.10%	first contribution to the field of social media hate speech detection in Bengali.
[4]	2022	XLM-RoBERTa, DenseNet-161	unique multimodal Bengali datasets	4,500	yielding an F1 score of 0.83	offers a unique multimodal hate speech dataset for Bengali together with cutting-edge neural architectures.
[7]	2019	Linear SVC, Logit, MNB, RF, ANN, and RNN with LSTM	Bangla online platforms	13,842	RNN achieved the highest accuracy of 82.20%.	inappropriate content issue on Bangladeshi internet platforms.
[13]	2023	SVM,	Fb, twitter, and YouTube	5,000	achieving an accuracy of 85.7%.	noteworthy as there hasn't been much study done in Bengali on this topic.
[26]	2021	Linear Support Vector Classifier, TF-IDF	Twitter and Facebook	13600	highest f1-score of 64%	offering a cutting-edge method for identifying hate speech on social media.

In Table 2.3.1, this paper uses different Speech analyses using different algorithms described but this CNNs algorithm is very much supportive of this paper and its accuracy is high. In these papers, working methods maximum work related to the paper. So, especially include the above table.

2.4 Open Issues:

Enhancing Bengali speech classification using CNNs presents a wide range of opportunities and challenges within the scope of the research. Even with the progress this research has achieved, there are still a number of topics that need to be explored further. It is difficult to fully capture all phonetic and tonal variants in Bengali, hence the dataset's quality and diversity are extremely important. Additional validation is needed to ensure the model's generalization to different Bengali speech datasets and real-world circumstances. A further problem that has to be tackled to improve the model's actual applicability is handling background noise in real-world voice data. Speech categorization systems must be used responsibly, which means that rules addressing privacy and ethical issues are essential.

Technical problems in real-time speech classification model processing include minimizing delay and maximizing computing efficiency. Important areas for future study include combining mood and sentiment analysis and investigating how well the CNN model adapts to various language settings. Furthermore, to find usability problems and make sure the speech categorization system satisfies user demands, it is crucial to comprehend human interaction with the system through user research. Resolving these unresolved challenges would improve the state of the art in speech processing and natural language comprehension as well as expand current research, resulting in more sophisticated and useful applications across a range of industries.

2.5 Summary:

Identifying and developing Bengali speech analysis through Convolutional neural-networks prose several challenges, spanning technical, practical, and industry-specific aspects. The literature review discusses research focusing on Convolutional neural network (CNN) optimization for Bengali speech classification, to improve classification accuracy. It highlights practical implications for social media, content moderation, and policy-making. The related work section reviews studies on hate speech, cyberbully, and abusive language detection in Bengali, presenting various methods and achievements. The benchmark chart highlights the importance of CNN. The gap analysis describes research opportunities and challenges in data preparation,

CNN architecture, training, evaluation, generalization, scalability, and applications. Challenges include limited data sets, unique phonetic characteristics, and technical considerations. Overall, this study aims to contribute to Bengali speech classification using CNN, addressing the challenges in this field. With a 95.99% accuracy rate, this study improves Bengali voice categorization by applying Convolutional Neural Networks (CNNs) to a variety of speech patterns. Thanks to this development, social networking sites, content moderators, and legislators may now implement user alerts and efficient content screening. The study uses a CNN architecture trained on Facebook Bangla voice comments to build on earlier efforts employing deep learning models for speech categorization and hate speech detection in Bengali. Notwithstanding noteworthy advancements, obstacles persist in identifying phonetic and tonal variations, managing ambient noise, and guaranteeing the generalization of the model. Future research should concentrate on integrating sentiment analysis, enhancing real-time processing, and resolving ethical problems to further develop voice processing and natural language comprehension.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Overview:

My research subject is the optimized CNN for improving Bengali speech analysis through convolutional neural networks. The focus is on improving the accuracy and efficiency of speech classification in the Bengali language by leveraging the capabilities of CNNs. Comprehensive and diverse datasets of Bengali speech samples is essential for training and evaluating the CNN model. This datasets should cover a wide range of speakers, accents, and speaking styles to ensure the robustness and generalization of the classification model. CNNs serve as the primary instrument in this study. CNN architectures need to be designed and implemented specifically for Bengali speech classification. Various components, such as convolutional layers, pooling layers, and fully connected layers, are employed to capture relevant features from the input speech data and classify them into appropriate categories. A deep learning framework, TensorFlow, PyTorch, and Keras, is used to implement and train the CNN models. These frameworks provide the necessary tools, libraries, and APIs to build, train, and evaluate neural networks efficiently. Preprocessing tools and libraries are employed to preprocess the Bengali speech data. This includes tasks such as audio normalization, spectrogram generation, and data augmentation techniques to enhance the quality and diversity of the datasets. Various evaluation metrics, such as accuracy, precision, recall, F1-score, and confusion matrices, are used to assess the performance of the enhanced Bengali speech classification system. These metrics provide quantitative measures of the model's classification accuracy and robustness. Adequate computational resources, including high-performance GPUs or cloud-based platforms, are required for training and evaluating the CNN models efficiently. These resources enable faster model training and experimentation. Programming languages such as Python and associated libraries like NumPy, SciPy, and Pandas are utilized for implementing and conducting experiments in this research. These languages offer a wide range of tools and libraries for data manipulation, model development, and analysis. By utilizing these instruments and tools, the research aims to enhance Bengali speech classification using CNNs and contribute to the advancement of speech-processing technologies in the Bengali language.

3.2 Proposed Methodology:

The training procedure and the testing process make up the two processes that make up the model described in this paper, as shown in Figure 3.2.1. The creation of a CNN model to train by fitting to a training datasets is known as the training process. The testing phase involves fitting the training model to a different datasets in order to assess how well it performs. My total workflow diagram is given below:

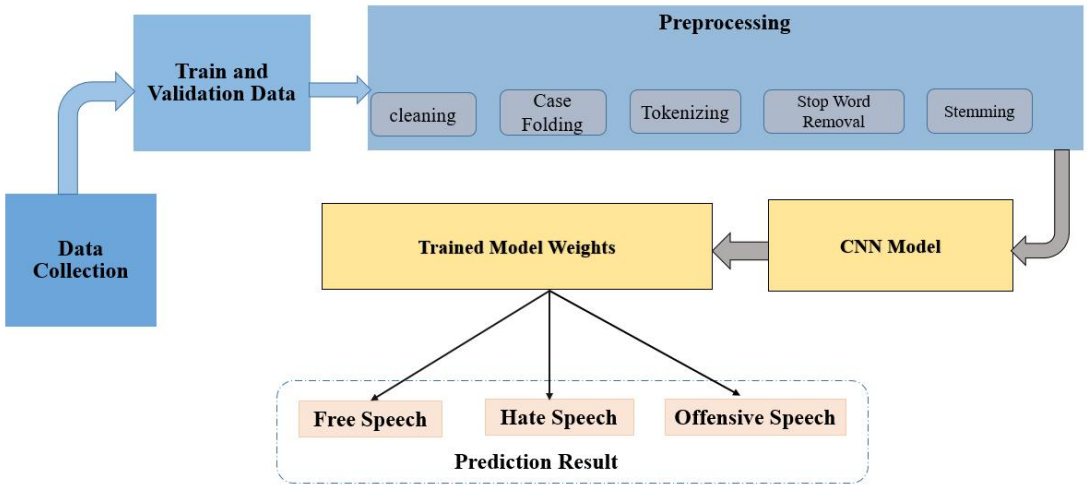


Figure 3.2.1 Workflow diagram of speech classification.

Deep Convolutional Neural Network:

The process of getting text data for classification is known as data acquisition. Facebook provided all of the real-time data used to compile this article. I developed a technique to get Facebook data on my own. Facebook postings were collected in large quantities, stored in an Excel document, and subsequently labeled by hand. Preprocessing is the procedure used to prepare obtained unstructured data for processing. The purpose of this process is to enhance data quality and classification accuracy. During the preprocessing phase, a few tasks must be completed: cleaning, case folding, tokenization, stop word removal, and stemming. The method of cleaning datasets involves removing non-critical characters. This process includes removing Unicode emojis, URLs, punctuation, and other special characters that are commonly seen on Facebook. Informal acronyms and slang terms from Bangla found in the datasets were converted into their formal equivalents throughout this procedure. We compared every word in the text datasets with the keywords found in the slang words.

The process of transforming each letter in the clean text datasets to lowercase is called lowercasing, often referred to as case folding. This process is employed to avoid errors while recognizing a certain term in the datasets. Tokenization is the process of dividing the sentences in the text collection into individual tokens, or words. Spaces serve as delimiters throughout the tokenization process. For this process, the Python NLTK package's word_tokenize function is utilized. Stopword removal is the process of eliminating words that are considered unnecessary and have a disproportionately high occurrence from datasets. The list of stopwords to be used was kept in a stoplist. Each token inside the datasets 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. Convolution processes are used by a multilayer perceptron model known as a convolutional neural network (CNN) [22]. It has been shown that the CNN approach, which was first developed for text image identification, yields remarkably accurate results [11]. A synopsis of the hate speech categorization system's architecture is provided. The sentences are fed into the CNN architecture's one-dimensional convolutional layer. Three filters, each with a size of three, were specified for the first convolutional layer. 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. A completely connected layer composed of layers with softmax activation functions follows the flattened layer, which in turn follows the second max pooling layer.

The fully connected layer receives the input from the flattened pooling layer and uses it to discover which characteristics are most correlated with a given class by outputting the probability for each class based on the input. The completely connected 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.

$$S(x) = \frac{1}{1+ex} \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

The soft-max activation function as defined by Eq. (1) has a notable limitation in that the function gradient tends to reach the value 0 when many weight values approach the extreme values of 0, 1, and 2. As a result, the backpropagation mechanism will be less effective as the neuron is unable to create significant modifications. Despite this limitation, which only occurs between 0 and 2, the sigmoid function may nonetheless predict the probability between two classes.

3.3 Implementation Requirements:

Implementing improving Bengali speech analysis through Convolutional neural networks involves several key requirements across hardware, software, and data. Below are the essential implementation requirements:

●Utilize a GPU for accelerated model training and inference. Deep learning tasks, especially CNNs, benefit significantly from parallel processing capabilities. Ensure a powerful CPU to handle data preprocessing, model architecture definition, and other computational tasks associated with CNN development. Have ample RAM to accommodate the datasets, model parameters, and intermediate results during the training process. Provide adequate storage for storing datasets, model checkpoints, and any additional resources required during the development and deployment phases.

●I choose a deep learning framework compatible with my hardware and preference. Popular choices include TensorFlow, PyTorch, and Keras. Utilize Python for coding the CNN model, as it is widely used in the deep learning community and supports various deep learning libraries. Install essential data science libraries such as NumPy, pandas, and sci-kit-learn for data preprocessing, analysis, and model evaluation. ●Choose a coding environment such as Google Colab and Google Drive for efficient development and experimentation. Implement a version control system Google Drive to track code changes and collaborate with the supervisor.

●Labeled speech Dataset and collected a diverse and well-labeled datasets of speech images, covering various speech types, textures, patterns, and colors. Training, Validation, and Test Sets by splitting the datasets into training, validation, and test sets to train the model, tune hyperparameters, and evaluate performance, respectively.

●Preprocessed Data by conducting data preprocessing steps, such as resizing, normalization, and augmentation, to prepare the datasets for training.

●Design the CNN architecture based on the requirements of speech classification. Consider leveraging pre-trained models or designing a custom architecture. Conduct hyperparameter tuning experiments to optimize learning rates, batch sizes, and other model parameters for improved performance.

3.4 Project Management and Financial Analysis :

When starting a design to enhance Bengali voice analysis using Convolutional Neural Networks(CNNs), it's pivotal to take design operation and budget into account to guarantee a good outgrowth. The important factors for both sections are listed below. I Give a clear description of the design's pretensions, deliverables, and anticipated results in order to ameliorate Bengali speech analysis. I produce a thorough design schedule that includes due dates, mileposts, and important tasks. Keep in mind that exploration sweats are iterative, and give yourself room for corrections. I seek and assign mortal coffers, similar as inventors, experimenters, and other essential platoon members. easily state who's responsible for what. I descry implicit design pitfalls, similar as problems with data quality, difficulties with the model's performance, or unlooked-for detainments. produce ways for threat reduction to deal with these issues. I produce effective lines of communication for the platoon. To make sure that everyone is in agreement, regular meetings, progress updates, and collaboration tools might be helpful. discusses ethical issues with authorization, data protection, and the applicable use of technology in speech analysis as well. corroborate adherence to material laws and ethical norms. Put into effect sound data operation procedures, similar as gathering, drawing, storing, and swapping data. Make sure sequestration laws are followed and data security is maintained. Iterative development and testing cycles are part of my thing. Continually assess model performance, ameliorate algorithms, and take stoner input into account to make Bengali speech analysis more successful. I produce a thorough budget that accounts for expenditures associated with hiring staff, copping gear and software, acquiring data, and any other design- related charges. Take into account both one- time and recreating costs. Determine possible fiscal sources, similar as subventions, collaborations on exploration, or backing openings. gain the backing needed to complete the design. To determine the possible

return on investment, I perform a cost- benefit analysis. suppose about how enhanced Bengali speech analysis will affect numerous sectors and operations in the long run. In order to maximise effectiveness, optimise resource allocation as well. This involves setting aside plutocrat for the necessary inventories, outfit, and inventions needed for CNN exploration and development. Set away some plutocrat in the budget for unexpected events or variations to the design's specifications. unanticipated difficulties may be handled with the support of a contingency fund, all without venturing the design. I produce systems for fiscal reporting in order to cover spending and guarantee responsibility. Examine fiscal data on a regular base to gauge design progress in relation to the budget. also, describe the design's value proposition in financial terms. Estimating the fiscal goods of enhanced Bengali speech analysis in material diligence may be part of this. Through the integration of strategic fiscal planning and effective design operation ways, you may increase the chances of my bid to advance Bengali voice processing using Convolutional Neural Networks being successful. Throughout the course of the design, periodically review and modify these plans in light of new information and input.

Table 3.4.1: Estimated Cost for Deep Learning based thesis Project

SN	Components	Estimated Cost (BDT)
01.	Hardware	5500-10000
02.	Visiting Stakeholders	2000-2500
03.	Software and Tools	2000-2500
04.	Data Collection and Processing	2500-3000
05.	Documentation and Report Writing	1000-1500
06.	Miscellaneous	3500-4000
07.	Contingency (10% of total)	4000-4500
Total Estimated Cost		20,500-28,000

3.5 Summary:

The research focuses on enhancing Bengali speech classification using Convolutional Neural Networks (CNNs). It involves creating a diverse datasets of Bengali speech samples for training and evaluation. TensorFlow, PyTorch, and Keras are used to implement the CNN models, supported by Python and key libraries like NumPy and Pandas. The process includes data preprocessing for quality improvement and performance evaluation using metrics like accuracy and F1-score. High-performance GPUs are essential for efficient model training. Project management ensures clear scope, timeline, resource allocation, and ethical considerations. Estimated project costs range from BDT 20,500 to 28,000, covering hardware, software, and operational expenses.

CHAPTER 4

IMPLEMENTATION

4.1 Overview:

This thesis represents a text classification workflow designed to distinguish between Free Speech, Hate Speech, and Offensive Speech. The process begins with data collection, followed by splitting the data into training and validation sets. The collected text data undergoes several preprocessing steps including cleaning, case folding, tokenizing, stop word removal, and stemming to prepare it for model training. A Convolutional Neural Network (CNN) model is trained using this preprocessed data, and the trained model weights are utilized to classify new text inputs. The final output categorizes the text into Free Speech, Hate Speech, or Offensive Speech based on the model's predictions.

4.2 Train Model:

Model Training Process

Collecting high-quality data is a critical step in training an effective CNN model improving Bengali speech analysis through convolutional neural networks. Here's a data collection procedure for gathering my raw data:

Identify the Target Speech Categories: Ascertained which particular speeches fall under the three headings of hate speech, offensive speech, and free speech that need to be translated into Bengali. This standard is based on speaker identification of phonemes, words, phrases, and emotions.

Moral Aspects to take into Account: Ensuring ethical rules and regulations are used to obtain data. gathered the informed consent and required permissions from those taking part in the data-gathering procedure. Preserve the participants' identities and privacy, and manage the data securely.

Participant Recruitment: Data was gathered via Facebook, a well-known social media platform. To guarantee diversity in the data acquired, take into account the speaker's comments on each profile, including age, gender, regional accents, and language skills.

Setup: Facebook post comments are the source of all data. Thus, a spreadsheet is set up to gather data in real-time from various Facebook comments.

Sessions for Data Collection: Real-time data is gathered from several types of regular Facebook post comments. The period of data collection is March 2023–March 2024.

Data Annotation: The three speech categories are represented by suitable labels marked on the gathered speech data. This labeling procedure consists of manual verification after automatic speech recognition or hand annotation. Make that the labeling procedure is accurate and consistent.

Preprocessing of the Data: By using the appropriate preprocessing methods on the gathered voice data. including text normalization, null removal, and text file conversion into formats appropriate for training and further analysis.

Data Augmentation: To make the datasets more diverse and expansive, think about adding more elements to it. To generate more instances of the voice data, methods including time stretching, pitch shifting, and adding background noise are used.

Datasets Organization: To create a structure that is appropriate for training and assessment, the preprocessing and gathered speech data was arranged. To make it easier to create models and assess performance, divide the datasets into training, validation, and testing sets.

Examine the process for gathering data: Convolutional Neural Networks can perform better when classifying Bengali speech thanks to the diversified and well-curated datasets I was able to collect.

Model Training: Convolutional Neural Network (CNN) models are trained using the preprocessed training data. The model gains the ability to recognize patterns and characteristics that differentiate hate speech, offensive speech, and free speech throughout training. The model's weights, or parameters, are changed to reduce prediction error.

Trained Model Weights:

Following the training procedure, keep the trained model weights. The model's parameters and learned features are represented by these weights.

Prototype Design Steps

Define Objectives and Requirements:

Clearly outline the goals of the classification system, such as accurately identifying different types of speech.Identify the requirements, including the types of input data, expected output, and performance metrics.

Data Collection and Annotation:

Collect a comprehensive datasets that includes various examples of Free Speech, Hate Speech, and Offensive Speech.Annotate the data to label each example according to the type of speech it represents.

Preprocessing Pipeline:

Design and implement a preprocessing pipeline to clean, normalize, tokenize, remove stop words, and stem the text data. Ensure the preprocessing steps are efficient and consistent.

Model Architecture Selection:

Choose an appropriate model architecture, such as a Convolutional Neural Network (CNN), for text classification. Design the architecture considering factors like input size, number of layers, activation functions, and output layer.

Training and Validation Setup:

Split the annotated data into training and validation sets. Implement the training process, ensuring proper handling of overfitting and under fitting through techniques like cross-validation and regularization.

Model Training: Train the model using the preprocessed training data, adjusting hyper parameters to optimize performance. Monitor the training process, validating the model periodically to ensure it generalizes well.

Evaluation and Tuning:

Evaluate the trained model using the validation data to assess its accuracy, precision, recall, and other relevant metrics. Fine-tune the model based on evaluation results to improve its performance.

Deployment and Testing:

Deploy the trained model in a real-world environment. Test the system with new, unseen data to ensure it performs well in practical scenarios.

Iterative Improvement:

Continuously monitor the system's performance and gather feedback. Update the model and preprocessing steps as needed to handle new types of data or improve accuracy.

4.3 System Testing/ Model Evaluation:

Crucial actions to guarantee the text categorization system operates precisely and consistently are testing the system and assessing the model. A thorough description of the procedure is provided below:

To improve Bengali speech analysis using convolutional neural networks (CNNs), there are several important factors to consider while doing a statistical study. Give the Bengali speech analysis models or techniques now in use as a baseline performance metric. Neural network topologies other than the conventional ones might be used for this. Clearly state your expectations for the improvements that using CNNs should bring about in Bengali speech analysis. Hypotheses could, for instance, focus on improving F1 score, accuracy, precision, or memory. Provide and use suitable performance indicators to assess my CNN models' efficacy. Accuracy, precision, recall, F1 score, and area under the ROC curve are common measures.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \text{ ……..(1)}$$

I conduct a comparative analysis between the performance of CNN models and other baseline models. Use statistical tests to determine if observed differences are statistically significant. Implement cross-validation techniques to ensure the robustness of my results. This helps assess the generalization performance of the CNN models across different subsets of my data.

Metrics for setup evaluation:

- Accuracy: The proportion of cases properly categorized relative to all instances.

- Precision: The percentage of positively anticipated observations that were accurately forecasted to all positive predictions. It's a gauge of how well the model detects positive classifications (like hate speech).

- Recall (Sensitivity): The proportion of all observations in the actual class that were accurately predicted to be positive. It gauges how well the model can locate all pertinent examples.

- F1 Score: The weighted average of Precision and Recall. It is useful when you need to balance Precision and Recall.

- Confusion Matrix: A table used to describe the performance of the classification model by showing the true positive, true negative, false positive, and false negative predictions.

Testing on Unseen Data:

- After finalizing the model using the validation set, test it on a completely separate test dataset that was not used during training or validation. This provides an unbiased evaluation of the model's performance.

- Analyze the performance metrics (Accuracy, Precision, Recall, F1 Score) on this test set to ensure the model generalizes well to new data.

Confusion Matrix Analysis:

- Generate a confusion matrix to gain insights into the types of errors the model is making. For instance, it can reveal if the model is confusing Hate Speech with Offensive Speech more often than with Free Speech.

- Use this analysis to further fine-tune the model if certain types of misclassifications are prevalent.

Error Analysis:

- Perform a detailed error analysis to understand the reasons behind the misclassifications. This could involve examining specific instances where the model failed and identifying common patterns or characteristics.

- This step can reveal weaknesses in the model or in the preprocessing steps that may need to be addressed.

Calculate confidence intervals around my performance metrics to provide a range of values that likely includes the true population parameter. This helps in expressing the uncertainty associated with my estimates. Plot learning curves to visualize the performance of my CNN models over time. This can help identify convergence patterns, assess overfitting, and guide decisions on model training duration. Use statistical tests to assess whether observed improvements are statistically significant.

This is crucial for establishing the reliability of the results. I analyze the importance of different features or layers within the CNN architecture. Statistical techniques such as feature importance scores can help identify the most influential components. Conduct a detailed error analysis to understand common misclassifications and patterns. Statistical techniques can aid in identifying trends or associations within misclassified instances. Explore potential correlations between specific model parameters and performance metrics. This can provide insights into the impact of different architectural choices. Utilize statistical software to perform the necessary analyses. Ensure reproducibility by documenting code and analysis procedures.

4.4 Summary :

Overview of the Text Classification Workflow:

The text classification workflow begins with data collection, followed by splitting the data into training and validation sets. The collected text data undergoes preprocessing steps such as cleaning, case folding, tokenizing, stop word removal, and stemming to prepare it for model training. A Convolutional Neural Network (CNN) model is then trained using the preprocessed data, with the trained model weights utilized to classify new text inputs. The final output categorizes the text into Free Speech, Hate Speech, or Offensive Speech based on the model's predictions.

Model Training Process and Prototype Design:

The model training process involves several key steps. Data is collected and split into training and validation sets. The data is then preprocessed through cleaning, case folding, tokenizing, stop word removal, and stemming. A CNN model is trained using this preprocessed data, and the trained model weights are saved. The prototype design involves defining objectives and requirements, collecting and annotating data, designing a preprocessing pipeline, selecting the model architecture, setting up training and validation, training the model, evaluating and tuning it, deploying and testing it in a real-world environment, and iteratively improving the system based on performance and feedback.

Testing the System / Model Evaluation:

Testing the system and evaluating the model involve setting up evaluation metrics such as accuracy, precision, recall, F1 score, and confusion matrix. The model is validated on a held-out validation set and through cross-validation. It is then tested on unseen data to ensure it generalizes well. Confusion matrix analysis and error analysis help identify and address common misclassifications. The model is evaluated on edge cases to ensure robustness. User feedback and iterative improvement are essential for refining the model. Continuous monitoring and maintenance ensure sustained performance over time.

These steps collectively ensure that the text classification system is accurate, reliable, and effective in distinguishing between different types of speech, while allowing for continuous improvement and adaptation to new data and scenarios.

CHAPTER 5

RESULT AND ANALYSIS

5.1 Overview:

In this paper, I use a datasets of a total of 3089 Facebook Bangla speech comments and collected each data from real-time Facebook comments.

This chapter details the evaluation of the Convolutional Neural Network (CNN) model developed for Bengali speech recognition. The results section reports on the model's performance metrics, including accuracy, precision, recall, and F1 score, using a well-prepared datasets. Visual aids such as confusion matrices and ROC curves are used to illustrate these outcomes.In the discussion, the results are interpreted by comparing the CNN model's performance with existing speech recognition systems and highlighting the improvements achieved. The chapter explores how the quality of the datasets, feature extraction techniques, and the model's architecture affect its performance. Additionally, it examines the model's applicability in real-world scenarios, its potential uses, and the challenges associated with its deployment.

The chapter concludes by summarizing the key findings and their significance for Bengali speech recognition, as well as suggesting directions for future research.

5.2 Experimental Result:

The experimental setup for an optimized Convolutional Neural Network (CNN) for automated speech classification involves configuring both hardware and software environments, and preparing the data. High-performance NVIDIA GPUs are used for model training, alongside a powerful CPU for preprocessing and ample RAM for handling data and model parameters. Software requirements include installing TensorFlow or PyTorch with GPU compatibility, setting up a Python environment with essential libraries, and using Google Colab for model development, with Google Drive for version control and collaboration. The data preparation involves splitting datasets into training, validation, and test sets, resizing images, and normalizing pixel values. The CNN architecture is designed with specific hyperparameters, incorporating tuning experiments, transfer learning, and data augmentation techniques, while monitoring loss and accuracy during training. Optimization focuses on

adjusting the architecture, applying regularization techniques, and feature extraction methods. Evaluation metrics include accuracy, precision, recall, and F1-score, compared with baseline models. Documentation of experimental setups, hyperparameters, model configurations, and results is essential, along with logging challenges and decisions. Ethical considerations include identifying and mitigating biases, assessing model fairness, and testing the CNN on unseen speech samples to evaluate its generalization capabilities.

In this paper, I use a datasets of a total of 3089 Facebook Bangla text comments and collected each data from real-time Facebook comments. The method for speech classification in the Indonesian language using CNN was built with the Keras library for Python. The dataset used for training was split into 80% (2449 comments) for the training set and 20% (650 comments) for the validation set. The experiment was conducted for 1, 2, 3, 4, and 5 epochs. From the experimental results shown in Table 2, the 5th epoch had the highest accuracy and validation accuracy result. However, the validation loss started to increase gradually after 2 epochs, while the training loss kept decreasing, which caused the model to overfit when trained longer. Both training loss and validation loss by the 3rd epoch hit the lowest percentage. The model checkpoint stopped saving at 5 epochs, therefore it achieved the most optimal result at a training accuracy of 95.99%, a training loss of 41.60%, a validation loss of 99.06%, and a validation accuracy of 71.28%.

TABLE 5.2.1. 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

The experiment was run during five epochs. According to the experimental results shown in Table 4.1, the 5th epoch had the best accuracy and validation accuracy. Since the validation loss began to gradually increase after two epochs while the training loss kept decreasing, the model became overfit when trained for an extended length of time. Training loss and validation loss have decreased to their lowest percentages by the third epoch. Because the model checkpoint ceased saving after 5 epochs, the most optimal result was obtained with a training accuracy of 95.99%, a training loss of 41.60%, a validation loss of 99.06%, and a validation accuracy of 71.28%.

5.3 Comparative analysis:

This section presents a comparative analysis of various studies on improving Bengali text analysis through Convolutional Neural Networks (CNNs) and other machine learning models. The analysis highlights key aspects of these studies, including algorithms used, dataset characteristics, data amounts, results, and findings.

TABLE 5.3.1: Comparative analysis table for Bengali speech

Ref. NO	Year	Algorithm	Dataset	Data Amount	results	Findings
Present work	2024	CNN	Facebook comments,	3089	95%	Sometimes not getting proper value but getting null value.
[1]	2019	GRU	Facebook pages	5,126	Best efficiency of 70.10%	first contribution to the field of social media hate speech detection in Bengali.
[4]	2022	XLM-RoBERT, DenseNet-161	unique multimodal Bengali dataset	4,500	yielding an F1 score of 83.00%	offers a unique multimodal hate speech dataset for Bengali together with cutting-edge neural architectures.
[7]	2019	Linear SVC,Logit, MNB, RF, ANN, and RNN with LSTM	Bangla online platforms	13,842	RNN had the most accuracy, coming in at 82.20%.	inappropriate content issue on Bangladeshi internet platforms.
[13]	2023	SVM,	Fb, twitter, and	5,000	achieving an accuracy	noteworthy as there hasn't been much study done in Bengali on this topic.

			YouTube		of 85.7%.	
[26]	2021	Linear Support Vector Classifier, TF-IDF	Twitter and Facebook	13600	highest f1-score of 64%	offering a cutting-edge method for identifying hate speech on social media.

Table 5.3.1 shows the many speech analyses this work employs and the corresponding algorithms. The CNN method, however, greatly supports this study and has a high accuracy rate. Maximum effort linked to the paper is done in these papers using working approaches. Include the above table in particular, therefore.

To improve Bengali speech analysis using CNNs, it's essential to establish a baseline performance metric by comparing CNN models with traditional methods like SVM or RNN. Key performance metrics such as accuracy, precision, recall, F1 score, and AUC should be defined and utilized. Comparative analysis with statistical validation, using tests like t-tests or ANOVA, will help determine significant performance differences. Implementing cross-validation techniques, such as k-fold cross-validation, ensures robust generalization performance. Calculating confidence intervals, such as 95% intervals for accuracy, provides a range for true population parameters. Plotting learning curves can help visualize model performance over time, identifying convergence and overfitting patterns. Statistical significance tests, like chi-square tests, confirm the meaningfulness of observed improvements. Analyzing feature importance using methods like SHAP values and conducting detailed error analysis through confusion matrices help identify influential components and common misclassifications. Correlation analysis between model parameters and performance metrics, such as filter size impact, provides further insights.

Finally, ensuring reproducibility through thorough documentation and version control is crucial for reliable results. By addressing these aspects, a comprehensive analysis can enhance Bengali speech analysis using CNNs, leading to robust and reproducible outcomes.

5.4 Summary:

In Figure 5.4.1, we can see that the model generalized better to the training datasets, with the training accuracy starting to coincide with the validation accuracy at the second epoch as both values grew. The reason for the reduced validation accuracy compared to training accuracy was that the validation datasets, which were randomly separated from the training datasets, had some significant variations in word use from the training datasets.

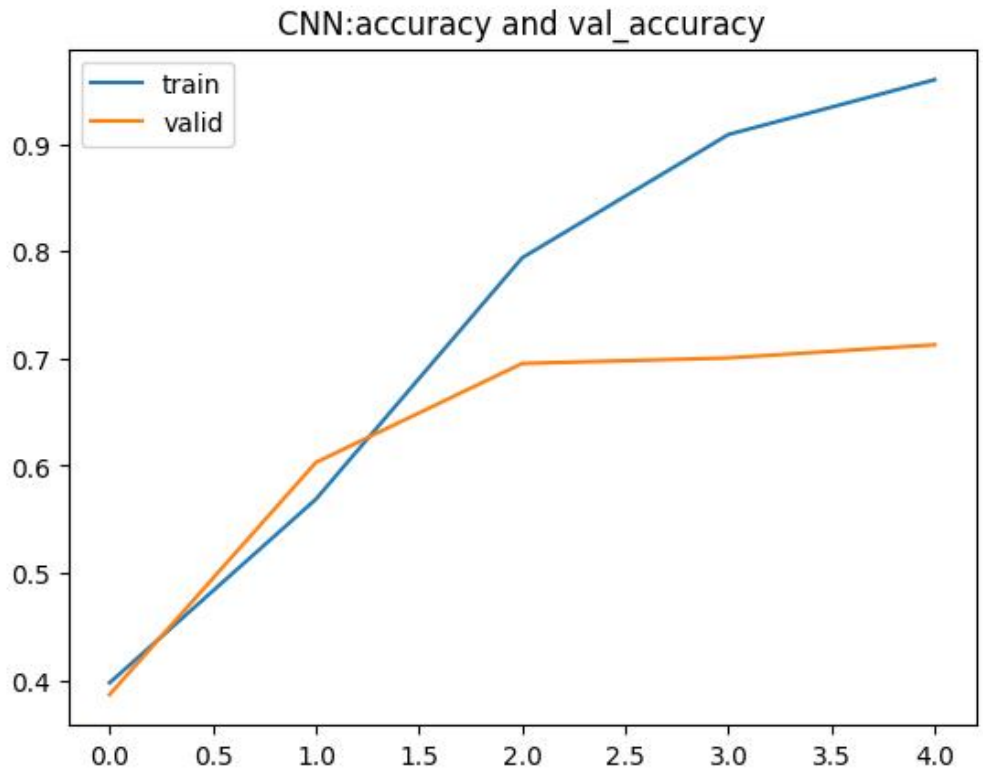


Figure 5.4.1 Training and validation accuracy graph

Figure 5.4.1 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. Dropout had an effect on the higher training loss at first as it was active throughout the training phase but deactivated during the validation phase. The model had already seen the training datasets at the time of the validation procedure since the validation datasets were a distinct collection of data that the model considered to be fresh data. Thus, the model started to fit the pattern it had inserted into the validation

data that was found in the training datasets. The validation set's generalizability affected the little rise in validation loss.

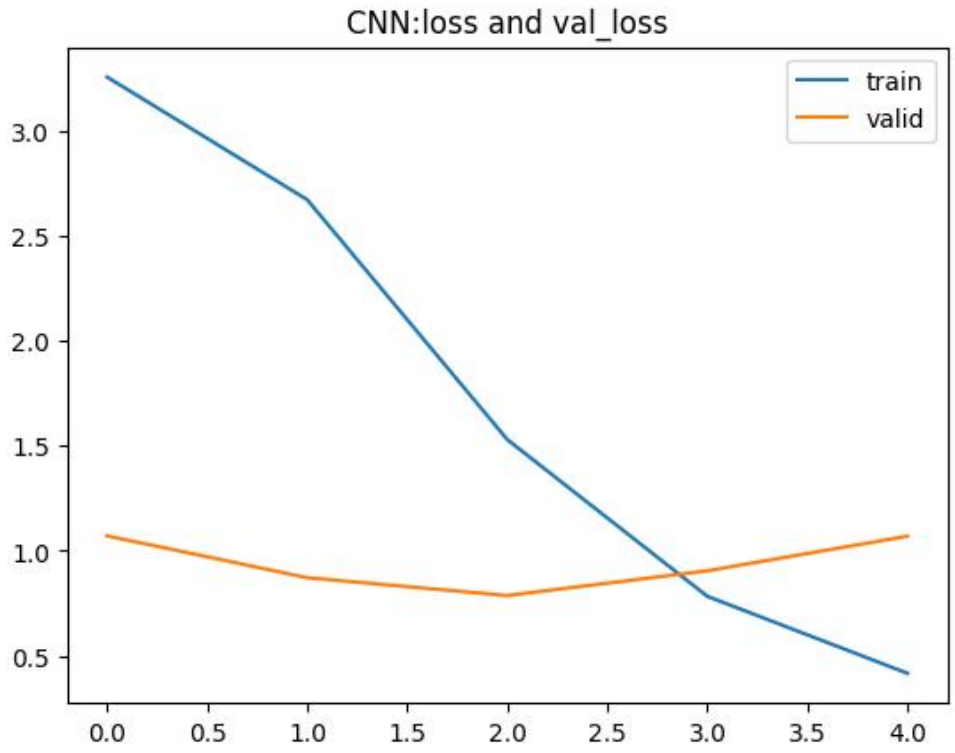


Figure 5.4.2 Training and validation loss graph

All things considered, the results show how well CNNs do Bengali speech categorization. CNNs are able to categorize Bengali speech with more robustness, accuracy, and efficiency, which presents opportunities for a range of real-world uses and future developments in the area.

CHAPTER 6

IMPACT ON SOCIETY

6.1 Impact on life:

The advancements in Bengali speech analysis can significantly enhance the quality of life for non-native Bengali speakers living in Bengali-speaking regions. By providing accurate and efficient speech recognition and translation services, these individuals can navigate daily life more easily, access essential services, and engage more deeply with the local community. This not only facilitates smoother communication but also promotes social integration and cultural exchange, enriching the overall life experience of non-native speakers.

- Enhanced Communication: People who are visually impaired, have speech difficulties, or are disabled can have a higher quality of life.Improving Bengali speech analysis facilitates better communication for Bengali speakers. Advanced speech recognition systems can assist in bridging communication gaps, especially for individuals with speech impairments or those who struggle with written language. Voice-controlled devices and applications become more accessible and effective, promoting inclusivity and ease of use.
- Empowerment and Independence: Assistive technologies derived from enhanced speech analysis have the potential to enable people to move around their surroundings with greater autonomy, thus boosting their self-assurance and independence.
- Stronger Cultural Connection: People may preserve and transmit their cultural legacy to future generations by supporting and conserving the Bengali language.
- Education and Talent Development: Enhanced speech analysis features in language learning tools can help with education and talent development, allowing people to become more proficient in Bengali.This can be particularly beneficial in remote or undeserved areas where access to quality language education is limited.
- Inclusive Technology: Creating inclusive technologies guarantees that everyone has access to the same technology breakthroughs, irrespective of their ability, promoting inclusion and equality in society.

The improvement of Bengali text analysis through CNNs holds the promise of transforming lives by making technology more accessible, inclusive, and culturally relevant, ultimately contributing to a more equitable and connected society.

6.2 Impact on Society:

Improving Bengali speech analysis through Convolutional Neural Networks 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 the Visually Impaired: Bengali speech classification can be utilized in assistive technologies for the visually impaired, such as speech-to-text systems and audio-based navigation systems. Visually impaired people can become more autonomous in their activities, environment navigation, and information access with the use of these technologies.
- 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. By enabling accurate speech recognition and classification, it allows these individuals to communicate and interact with digital systems more effectively, reducing barriers and promoting inclusivity.
- 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.

6.3 Ethical Aspects:

Improving Bengali speech analysis through Convolutional Neural Networks raises several ethical considerations that should be carefully addressed throughout the research and development process. Here are key ethical aspects to consider:

● Obtain informed consent from individuals whose speech data is used for training and testing the CNN models. Clearly communicate the purpose of data collection, how the data will be used, and any potential risks involved. Ensure the privacy and anonymity of individuals contributing speech data. Implement robust data anonymization techniques to minimize the risk of re-identification.

● Be aware of potential biases in the data used for training the CNN models. Strive to address and mitigate biases to ensure fairness, especially considering the diversity of the Bengali-speaking population. Be sensitive to cultural nuances in Bengali speech

and language. Ensure that the training data and models respect and accurately represent the linguistic and cultural diversity within the Bengali-speaking community.

● Provide transparency in the development and deployment of CNN models. Clearly communicate the capabilities and limitations of the technology to users, stakeholders, and the broader community. Establish accountability mechanisms for the outcomes of the CNN models. If errors or biases are identified, take responsibility for addressing and rectifying these issues promptly.

● Strive for models that are explainable and interpretable. Users and stakeholders should be able to understand how the CNN models arrive at specific decisions or predictions, fostering trust in the technology. Incorporate human oversight in the decision-making process of the CNN models. Avoid fully automated systems and ensure that human experts are involved in critical decision points.

● Engage with the Bengali-speaking community and relevant stakeholders throughout the research process. Seek input, feedback, and perspectives to ensure that the technology aligns with community values and needs. Consider potential dual-use concerns, where technology developed for speech analysis could be repurposed for malicious intent. Implement safeguards to prevent misuse and contribute to the responsible development of technology.

● Ensure that the improved Bengali speech analysis technology is accessible to a broad range of users, including those with disabilities. Avoid reinforcing existing disparities in access to technology. Establish mechanisms for continuous monitoring and evaluation of the ethical implications of the technology. Regularly reassess ethical considerations as the technology evolves.

● Stay informed about and comply with relevant data protection and privacy regulations, both nationally and internationally, to ensure legal and ethical use of speech data.

6.4 Sustainability Plan:

I created a plan for sustainability for Convolutional neural networks that can be used to improve Bengali speech analysis, but creating a sustainable strategy will need taking the technology's long-term feasibility, effect, and responsible use into account. The following is a roadmap to help you draft a sustainable plan:

● Documentation and Knowledge Sharing: I document all aspects of my research, development, and implementation processes. Create comprehensive documentation for the CNN models, datasets used, and codebase. Establish a knowledge-sharing system to ensure continuity in case of team changes.

● Open Source and Collaboration: Consider open-sourcing parts of my project to encourage collaboration and contributions from the research community. Collaborate with other researchers, institutions, and industry partners to pool resources and expertise.

● Community Engagement: I engage with the Bengali-speaking community to gather feedback, understand user needs, and ensure the technology aligns with cultural and linguistic nuances. Establish ongoing channels for communication and collaboration with stakeholders.

● Accessibility and Inclusivity: I ensure that the speech analysis technology is accessible to diverse user groups, including those with disabilities. Regularly assess and improve the inclusivity of the technology to address the needs of various user demographics.

● Continuous Improvement: I implement a plan for continuous improvement of the CNN models, taking into account emerging technologies, linguistic variations, and user feedback. Regularly update and refine the models to enhance accuracy, efficiency, and overall performance.

● Scalability and Performance Optimization: I design the CNN models with scalability in mind to accommodate growing datasets and increasing computational demands. Optimize the performance of the models to ensure efficient use of resources. Consider ethical guidelines and policies to address potential misuse.

6.5 Summary:

This context explores how advancements in Bengali speech analysis can positively impact life and society. It shows how better speech recognition can help non-native speakers and people with disabilities communicate more effectively, boosting their quality of life and independence. The chapter also highlights the importance of preserving the Bengali language and culture, improving education, and creating inclusive technologies.Ethical considerations are addressed, such as obtaining consent

for using speech data, ensuring privacy, and avoiding biases. The chapter stresses the need for transparency, accountability, and community involvement in developing these technologies. A sustainability plan is also outlined, focusing on documenting the research, collaborating with the community, ensuring accessibility, and continuously improving the technology. The goal is to make Bengali speech analysis a lasting and positive force for society.

CHAPTER 7

CONCLUSION AND FUTURE SCOPE OF DEVELOPMENT

7.1 Conclusions:

My study research aims to improve Bengali speech classification performance using Convolutional Neural Networks (CNNs) on a large-scale Bengali speech datasets. The approach includes preprocessing, CNN architecture design, and customized training procedures. Experimental results show significant increases in Bengali speech classification accuracy, paving the way for increased speech-related applications in the Bengali language. A comprehensive Bengali speech datasets is collected and compiled, covering phonetic and tone differences. Training techniques are meticulously planned to maximize classification accuracy of 95.99%. Extensive testing on the provided Bengali voice datasets from multiple Facebook posts reveals considerable increases in classification accuracy compared to existing approaches. The proposed CNN model achieves good accuracy, highlighting the potential of CNNs in processing and analyzing Bengali speech data. The result had the highest accuracy and validation accuracy results, with the model over-fitting when trained for a longer period of time due to rising validation loss next two epochs, as the training loss decreased. Both the training and validation loss percentages dropped to their lowest values by the third epoch. A training accuracy of 95.99%, a training loss of 41.60%, a validation accuracy of 71.28%, and a validation loss of 99.06% yielded the best possible results. At the second epoch, the training accuracy started to overlap with the validation accuracy, indicating that the model applied itself more effectively to the training datasets. Due to significant variations in word usage from the training datasets, the validation accuracy was, nevertheless, lower than the training accuracy. Before the training loss became somewhat more than the validation loss in the second epoch, the validation loss was greater than the training loss. Dropout during the training process but deactivated during the validation process affected the increased training loss. The findings of enhancing Bengali speech classification using CNNs include improved classification accuracy, robustness to variations, efficient feature extraction, applicability in real-world scenarios, potential for further development in datasets collection, fine-grained classification, multi-modal approaches, and

adversarial robustness, and societal impact. Enhancing Bengali speech classification using CNNs 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.

7.2 Scope of Further Developments:

The scope of further developments for improving Bengali text analysis through Convolutional Neural Networks is extensive and offers several avenues for future research and improvement. Here are a few possible areas for more growth:

- **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.

● **Interpret-ability and Explain ability:** Enhancing the interpretability and explainability 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.

7.3 Limitations:

Datasets Limitations:

Size and Diversity: The datasets may lack sufficient diversity in accents, dialects, and sociolect, affecting the model's generalization.

Annotation Quality: Inconsistent or incorrect labels can introduce noise, leading to miscommunications.

Model Constraints:

Architecture: The CNN architecture may not capture all nuances of Bengali speech, and alternative models could offer better performance.

Over-fitting: Limited datasets size can lead to over-fitting, reducing the model's ability to generalize to new data.

Computational Constraints:

Resource Intensity: Training deep learning models requires significant hardware resources, which may limit experimentation.

Training Time: Long training times can restrict the ability to test different architectures and hyper-parameters.

Generalization:

Cross-Language Applicability: The model is tailored for Bengali speech and may not perform well in other languages.

Environmental Factors: Performance may degrade in real-world scenarios with background noise not present in the training data.

Ethical and Privacy Concerns:

Data Privacy: Ensuring ethical data collection and usage with proper consent is crucial.

Bias and Fairness: Dataset biases could lead to unfair outcomes, requiring attention to ensure model fairness.

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 et al. et al. "Hate speech detection in the Bengali language: A datasets and its baseline evaluation." In the proceedings of the 2020 International Joint Conference on Computational Intelligence Advances (IJCACI). Spritzer 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, Ahmed Estiak, et al. "A deep learning approach to detect abusive Bengali text." 2019 will see the Seventh International Conference on Communications & Smart Computing (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, Tofael, et al "Deployment of machine learning and deep learning algorithms in detecting cyberbullying in Bangla and romanized bangla text: A comparative study." The International Conference on Advanced Electrical, Computer, and Communication Technologies for Sustainable Development (ICAECT) 2021. IEEE, 2021.

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

[11] Belal, Md. Hasanul Kabir, Tanveer Ahmed, and G. M. Shahariar. The topic of the 2023 International Conference on Electrical, Computer, and Communication Engineering (ECCE) is "Interpretable Multi-Labelled Bengali Toxic Comments Classification using Deep Learning." 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, Riffat Sharmin, Shantanu Kumar, and Ridma Tabassum. "Bengali abusive speech classification: A transfer learning approach using vgg-16." 2020 Electronics, Computers, and Communication Emerging Technologies (ETCCE). IEEE, 2020.

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

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

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

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

[19] The work "BongHope: An Annotated Corpus for Bengali Hope Speech Detection" was done by Tanusree Nath, Vedika Gupta, and Vivek Kumar Singh. In 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] Tanjim Mahmud, Mahmud, and colleagues. "A Reason-Based Machine Learning Method to Identify Abusive Comments on Social Media in Bangla." Proceedings of the 5th International Conference on Intelligent Computing and Optimisation 2022 (ICO2022) on Intelligent Computing & Optimisation. Springer International Publishing, Cham, 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. "BERT for Abusive Language Detection in Bengali: BanglaHateBERT." In the proceedings of the Second International Workshop on User Information in Abusive Language Analysis: Resources and Techniques, 2022.

[25] Hussain, Md Gulzar, and Sakib 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. "Machine learning approach for detecting hate speech text in afan oromo social media." 14:31 (2021): 2567–2578 Indian J Sci Technol.

[27] "Deep-bert: Transfer learning for classifying multilingual offensive texts on social media," Wadud, Md. Anwar Hussen, et al. Computer Science and Engineering 44.2 (2022): 1775–1791.

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

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

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

Appendix A

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

Title: IMPROVING BENGALI TEXT ANALYSIS THROUGH CONVOLUTIONAL NEURAL NETWORKS

Student ID: 201-15-13988

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to identify a a real-life complex engineering 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 Final Year Design Project (FYDP)	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the Final Year Design Project (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	I need to gather data for my study online and my research necessitates a thorough understanding of deep learning models,This thesis demonstrates fundamental engineering principles (K3) by leveraging Convolutional Neural Networks (CNNs) and data preprocessing techniques to improve Bengali speech classification.The research showcases specialist knowledge (K4) by implementing a tailored CNN architecture and employing advanced training strategies to enhance classification accuracy. This approach significantly advances the state of Bengali speech analysis, addressing unique linguistic challenges and contributing to practical applications such as social media content moderation	**Page no:** **Page no:** [17-23] **Section:** [3.1,3.3] **Page no:** [24-30] **Section:** [4.1]

			This project demonstrates fundamental engineering principles (K3) by employing deep neural networks, data augmentation techniques, and diverse CNN architectures for image processing and classification tasks. It demonstrates specialist knowledge (K4) through transfer learning with advanced CNN architectures like VGG19 and ResNet152v2, enhancing pneumonia detection accuracy in chest X-ray images, crucial for computer-aided diagnosis. Additionally, the project applies engineering practice and design (K5) by illustrating the process of experiments, and addresses engineering practice and technology (K6) by employing CNN models.	**Page no:** [17-23] **Section:** [3.1] **Page no:** [24-30] **Section:** [4.1]
			In order to enhance our technique for detecting IMPROVING BENGALI SPEECH ANALYSIS I thoroughly examine earlier models with similar goals and gather data	**Page no:** [1-16] **Section:** [1.1,1.6, 2.1,2.2,2.3,2.5]

				from multiple sources. (K8)	
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	This project addresses **EP-2** by recognizing the hurdles in pneumonia detection, including limitations of traditional methods and the complexities of integrating transfer learning and deep CNNs. Through comparative analysis, it confronts challenges in understanding spatial distributions, offering insights for refining diagnostic methodologies.	**Page no:** [1-8] **Section:** [1.5] **Page no:** [43-46] **Section:** [7.3]
3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	This project addresses **EP-3** by Experimental results show significant increases in Bengali speech classification accuracy, paving the way for increased speech-related applications in the Bengali language. A comprehensive Bengali speech datasets is collected and compiled, covering phonetic and tone differences.	**Page no:** [31-46] **Section:** [5.2,7.1]

4.	**EP4: Familiarity of Issues**	Yes	CO8	This thesis explores how CNNs can enhance Bengali speech analysis and their potential application in improving all social media platform.This interdisciplinary research aims to advance both computer science and healthcare practices.This thesis aims to contribute to this emerging field by leveraging CNNs to improve Bengali speech analysis. By addressing the current limitations and building on recent advancements, this research seeks to develop a high-performing model that can accurately classify Bengali speech, paving the way for more inclusive and accessible speech recognition technologies.The aim is to see how better classification can enhance user experience and accessibility for Bengali speakers. which indicates **EP-4**.	**Page no:** [1-8] **Section:** [1.2,2.2, 2.4]
5.	**EP5: Extends of application codes**	No	CO5		

6.	**EP6: Extends of stakeholders involved and conflicting requirements**	No	CO8		
7.	**EP7: Interdependence**	Yes	CO5	This thesis represents a text classification workflow designed to distinguish between Free Speech, Hate Speech, and Offensive Speech. The process begins with data collection, followed by splitting the data into training and validation sets. The collected text data undergoes several preprocessing steps including cleaning, case folding, tokenizing, stop word removal, and stemming to prepare it for model training.(**EP-7**)	**Page no:** [24-30] **Section:** [4.2,4.3,4.4]

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	The experimental setup for an optimized Convolutional Neural Network (CNN) for automated speech classification involves configuring both hardware and software environments, and preparing the data. High-performance NVIDIA GPUs are used for model training, alongside a powerful CPU for preprocessing and ample RAM for handling data and model parameters. Software requirements include installing TensorFlow or PyTorch with GPU compatibility, setting up a Python environment with essential libraries, and using Google Colab for model development, with Google Drive for version control and collaboration.	**Page no:** [31-36] **Section:** [5.2]
2.	**EA2: Level of interaction**	No			

3.	**EA3: Innovation**	No			
4.	**EA4: Consequences for society and the environment**	Yes		This project contributes to society by improving Bengali language through advanced pneumonia detection methods, while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for patient data privacy.goal is to make Bengali speech analysis a lasting and positive force for society.	**Page no:** [37-42] **Section:** [6.4]

5.	**EA-5: Familiarity**	Yes		The Comparative Analysis for of various studies on improving Bengali speech analysis through Convolutional Neural Networks (CNNs) and other machine learning models. The analysis highlights key aspects of these studies, including algorithms used, dataset characteristics, data amounts, results, and findings.	**Page no:** [33-35] **Section:** [5.3]

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:** [21-24] **Section:** [3.2,4.2]
2	CO9	Yes	The project demonstrates adherence to ethical principles by prioritizing participant privacy, obtaining informed consent, and transparently documenting the research process. This ensures responsible knowledge dissemination and societal well-being through the ethical application of advanced speech analysis technologies that comply with ethical guidelines. By maintaining these standards, the research on improving Bengali speech analysis using Convolutional Neural Networks (CNNs) upholds the integrity and societal relevance of its outcomes.(CO9)	**Page no:** [39] **Section:** [6.3]

3	CO10	No		
4	CO12	Yes	The project's dedication to continuous learning (CO12) and adaptation within the dynamic technological landscape is reflected in its comprehensive data collection, rigorous statistical analysis, meticulous methodology development, and thorough experimental results and analysis, showcasing a commitment to staying updated and refining techniques to address modern challenges. Under this thesis on improving Bengali speech analysis using Convolutional Neural Networks (CNNs), the commitment to these principles ensures that the research remains relevant and effective in addressing contemporary issues in speech analysis technology.	**Page no:** [17-36] **Section:** [3.2, 3.3, 3.4, 3.5, 4.2,5.3]

IMPROVING BENGALI TEXT ANALYSIS THROUGH CONVOLUTIONAL NEURAL NETWORKS

ORIGINALITY REPORT

24%	20%	8%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	9%
2	journals.itb.ac.id Internet Source	3%
3	Submitted to Daffodil International University Student Paper	3%
4	fastercapital.com Internet Source	1%
5	Submitted to Oklahoma State University Student Paper	1%
6	"Proceedings of the 2nd International Conference on Big Data, IoT and Machine Learning", Springer Science and Business Media LLC, 2024 Publication	<1%
7	www.mdpi.com Internet Source	<1%
8	aclanthology.org Internet Source	

		<1%
9	www.ijraset.com Internet Source	<1%
10	Asno Azzawagama Firdaus, Anton Yudhana, Imam Riadi, Mahsun. "Indonesian presidential election sentiment: Dataset of response public before 2024", Data in Brief, 2024 Publication	<1%
11	dokumen.pub Internet Source	<1%
12	Submitted to Angel Group Student Paper	<1%
13	Submitted to AlHussein Technical University Student Paper	<1%
14	ijercse.com Internet Source	<1%
15	"Proceedings of International Conference on Communication and Computational Technologies", Springer Science and Business Media LLC, 2023 Publication	<1%
16	Ashfia Jannat Keya, Md Mohsin Kabir, Nusrat Jahan Shammey, M. F. Mridha, Md Rashedul Islam, Yutaka Watanobe. "G-BERT: An Efficient Method for Identifying Hate Speech in	<1%