TELECOMMUNICATION FRAUD DETECTION BY ANALYZING THE CONTENT OF A BANGLA PHONE CALL

BY

Safiqul Islam Uzzal

ID: 203-15-14475

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

Sharun Akter Khushbu
Lecturer (Senior Scale)
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Md. Sazzadur Ahamed
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
July 2024

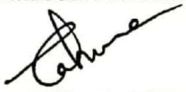
APPROVAL

This Project titled **"Telecommunication Fraud Detection By Analyzing the Content of a Bangla Phone Call"**, submitted by **Safiqul Islam Uzzal** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **13-07-2024.**

BOARD OF EXAMINERS

Dr. Md. Ismail Jabiullah (MIJ) — **Board Chairman**
Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Mr. Abdus Sattar (AS) — **Internal Examiner 1**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University



Ms. Zahura Zaman (ZZ) — **Internal Examiner 2**
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

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

DECLARATION

I hereby declare that this project has been done by me under the supervision of **Sharun Akter Khushbu, Lecturer (Senior Scale), Department of Computer Science and Engineering, Daffodil International University**. I also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

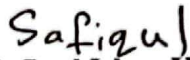
Supervised by:

Sharun Akter Khushbu
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Co-Supervised by:

Md. Sazzadur Ahamed
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:



Safiqul Islam Uzzal
ID: 203-15-14475
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

I am grateful and wish my profound indebtedness to **Sharun Akter Khushbu**, **Lecturer (Senior Scale)**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of **"*Natural Language Processing*"** to carry out this project. Her 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 **Head of the Department of CSE**, for his kind help in finishing my project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

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

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

ABSTRACT

Telecommunication fraud, especially through techniques like phishing calls, poses a significant threat, resulting in substantial financial losses and compromising sensitive personal data. While technological advancements have made life easier for consumers, they have also introduced new vulnerabilities. Phishing through phone calls, wherein individuals are tricked into revealing sensitive information, is a prevalent method employed by fraudsters. Traditional approaches use blacklists of known fraudulent numbers to detect phishing calls that are ineffective against new or previously unseen numbers. To tackle the problem, I propose a system leveraging conversation transcript analysis, combining an Automatic Speech Recognition (ASR) model to convert audio to text and Natural Language Processing (NLP) techniques to analyze the text. Two datasets are used in my experiment: a transcription dataset generated by a Bangla ASR system and a clean conversational dataset. I demonstrate that fine-tuning deep learning models to increase accuracy works well on transcription data provided by ASR. The study utilizes four deep learning models: Artificial Neural Network (ANN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Bidirectional Long Short-Term Memory (BiLSTM). The results demonstrate that the RNN model attains the best accuracy of 97% on the clean conversational dataset. In the transcription dataset, both LSTM and ANN models achieve highest accuracy of 83%. The results emphasize the capability of optimized deep learning models to improve the identification of fraudulent phone calls, especially in situations where the data is provided by ASR system. This provides a strong and effective approach to detect telecommunication fraud.

TABLE OF CONTENTS

Contents	Page No
Board of Examiners	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
CHAPTER 1: INTRODUCTION	**1-6**
1.1 Overview	1
1.2 Background and Present State	1-2
1.3 Problem Statement	2-3
1.4 Objectives	3
1.5 Scope and Limitations	3-4
1.6 Report Organization	4-5
1.7 Summary	5-6
CHAPTER 2: LITERATURE REVIEW	**7-12**
2.1 Overview	7
2.2 Related Works	7-11
2.3 Comparison between existing works	11-12
2.4 Open Issues	12
2.5 Summary	12
CHAPTER 3: METHODOLOGY/ REQUIREMENT ANALYSIS & DESIGN SPECIFICATION	**13-24**
3.1 Overview	13
3.2 Proposed Methodology	13-21
3.3 Hardware/ Software Requirement	21-22
3.4 Project Management and Financial Analysis	22-23
3.5 Summary	23-24

CHAPTER 4: IMPLEMENTATION **25-32**

4.1 Overview 25
4.2 Train Model 25-28
4.3 Model Evaluation 28-32
4.4 Summary 32

CHAPTER 5: RESULT AND ANALYSIS **33-38**

5.1 Overview 33
5.2 Experimental Result 33-36
5.3 Comparative Analysis 36-38
5.4 Summary 38

CHAPTER 6: IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY **39-40**

6.1 Impact on Life 39
6.2 Impact on Society & Environment 39
6.3 Ethical Aspects 39-40
6.4 Sustainability Plan 40
6.5 Summary 40

CHAPTER 7: CONCLUSION AND FUTURE WORK **41-42**

7.1 Conclusions 41
7.2 Further Suggested Works 41
7.3 Limitations 41-42

REFERENCES **43-45**

APPENDIX A **46-52**

LIST OF TABLES

TABLES	PAGE NO
Table 2.3.1: Comparative analysis of some work for Fraudulent Phone Call Detection	11-12
Table 3.2.1.1: Table for some Clean Conversations from my collected Dataset	15
Table 3.2.1.2: Table for some transcribed Conversations from my collected Dataset	15
Table 3.2.2.1: Comparison between Real & Augmented Conversation from my Dataset	16
Table 3.2.3.1: Comparison between Original & Clean Corpus from my dataset	17
Table 3.4.2.1: Estimated cost for this work	23
Table 4.3.1.1: Training Accuracy vs Validation Accuracy of deep learning models	29
Table 5.2.1.1: Training vs Testing Accuracy for my Conversational Dataset	34
Table 5.2.1.2: Precision, Recall & F1-Score for my Conversational Dataset	34
Table 5.2.2.1: Training vs Testing Accuracy for my Transcription Dataset	35
Table 5.2.2.2: Precision, Recall & F1-Score for my Transcription Dataset	36
Table 5.3.1: Accuracy Before vs After Fine-Tuning	37
Table 5.3.2: Performance Evaluation of Before vs After Fine-Tuning	37

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.2.1: Proposed Methodology	14
Figure 3.2.5.1: Architecture of ANN	19
Figure 3.2.5.2: Architecture of RNN	19
Figure 3.2.5.3: Architecture of LSTM	20
Figure 3.2.5.4: Architecture of BiLSTM	21
Figure 4.2.1.1: Working Flow of ANN for my work	26
Figure 4.2.1.2: Working Flow of RNN for my work	26
Figure 4.2.1.3: Working Flow of LSTM for my work	27
Figure 4.2.1.4: Working Flow of BiLSTM for my work	27
Figure 4.3.1.1: Training & validation accuracy and loss over the epochs of ANN on the clean conversational dataset	29
Figure 4.3.1.2: Training & validation accuracy and loss over the epochs of RNN on the clean conversational dataset	29
Figure 4.3.1.3: Training & validation accuracy and loss over the epochs of LSTM on the clean conversational dataset.	30
Figure 4.3.1.4: Training & validation accuracy and loss over the epochs of BiLSTM on the clean conversational dataset	30
Figure 4.3.2.1: Training & validation accuracy and loss over the epochs of ANN on the transcribed dataset	31
Figure 4.3.2.2: Training & validation accuracy and loss over the epochs of RNN on the transcribed dataset	31
Figure 4.3.2.3: Training & validation accuracy and loss over the epochs of LSTM on the transcribed dataset	31
Figure 4.3.2.4: Training & validation accuracy and loss over the epochs of BiLSTM on the transcribed dataset	32

Figure 5.2.1.1: Confusion Matrix of Deep Learning Models for the clean conversational test dataset	35
Figure 5.2.2.1: Confusion Matrix of Fine-tuned Deep Learning Models for the transcribed test dataset	36
Figure 5.3.1 Web Application Interface 1	37
Figure 5.3.2 Web Application Interface 2	38

LIST OF ACRONYMS

ML	Machine Learning
AI	Artificial Intelligence
ANN	Artificial Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
BiLSTM	Bidirectional Long Short-Term Memory
ASR	Automatic Speech Recognition
NLP	Natural Language Processing
MFS	Mobile Financial Service
WER	Word Error Rate
BNLP	Bangla Natural Language Processing

CHAPTER 1
INTRODUCTION

1.1 Overview

Concerns about fraud extend globally, and the telecommunications and mobile banking sector is no exception. Telecom fraud is an alarming issue because it can lead to significant revenue loss and can cripple the credibility and efficiency of telcos. Bangladesh has recently fallen in a trend of mobile banking setStatus As per [20], The number of banks that currently provide Mobile Financial Services (MFS) in Bangladesh is 17 and the average daily transaction is 4175.66 crores in January 2024 which shows clearly how this field is on the point of revolution from the perspective of Bangladesh. This growth has also spawned fraud trends such as impersonating real customer service representatives to scam people who are caught unawares. A significant proportion of the population is skeptical about using mobile financial services (MFS) over fear of falling victim to fraud in numerous forms. According to a report, compromised PINs were one of several methods of fraudsters who employed this. Furthermore, fraudsters regularly utilize impersonation and account takeovers. Oftentimes times scammers will call acting like reputable agents or managers. These are made up of many different types of manipulative practices to get victims to disclose personal information. With the recent growth of mobile financial services, it is imperative to reduce fraud and support the healthy development of such services by conducting research on fraudulent activities in mobile financial services in Bangladesh.

NLP models are already very successful in different realms, but the application of these models in this specific field was largely unexplored till now. I strongly believe that NLP models are best suited for detecting suspicious phone calls, based on language analysis. These conversational cues help these models analyze the language patterns that are written by the scammers which can assist in identifying possible attempts of scam/fraud. Integrating NLP technology in MFS and telecommunication can lead to significantly fewer people becoming a victim of fraud as well, increasing the security and trust of mobile financial services.

1.2 Background and Present State

Researchers in the field of detecting fraudulent phone calls have typically depended on two primary methods. An effective method is utilizing Call Detail Records (CDR) databases to detect and pinpoint instances of fraudulent behavior. The CDR files provide comprehensive data, including phone numbers, call durations, times, locations, etc. This information is helpful for analyzing calling trends and identifying any unusual activity that may suggest fraudulent behavior. Nevertheless, this approach

has constraints, particularly in the case of unfamiliar or previously unrecognized numbers, which pose difficulties in detecting intricate fraudulent schemes. Recently, there has been a growing inclination towards directly analyzing the content of phone calls. This novel approach entails converting audio conversations into written text and then utilizing NLP tools to examine the transcriptions. Through the analysis of the language content of phone conversations, researchers are able to identify minor signs and trends that might potentially suggest fraudulent actions. This technique has demonstrated considerable potential in enhancing the precision of fraud detection by examining the actual conversation content, rather than relying just on the information supplied by CDRs. In an effort to improve this approach's efficacy, scholars have investigated a range of deep learning and machine learning models. These models have been utilized to analyze the textual data gathered from transcriptions in order to detect fraudulent behaviors by examining linguistic trends. Furthermore, several researchers have taken a further step by constructing Android applications that use NLP models in real time to identify fraudulent activities during phone conversations. These programs have the capability to promptly notify users, hence improving the ability to detect and prevent fraudulent acts in real time.

I have investigated well-known deep learning models in my study, including BiLSTM, LSTM, RNN, and ANN, all of which have demonstrated efficacy in NLP classification tasks. Furthermore, I have experimented with the partial fine-tuning[21] method of these models and the outcomes are reported in this study. To assess and improve the effectiveness of these models in identifying fraudulent phone calls based on their textual content, important ideas including confusion matrices, classification reports, and hyperparameter tweaking have also been covered.

1.3 Problem Statement

Conventional telecom fraud detection systems mostly depend on blacklists of known fraudulent phone numbers, which prove useless when dealing with unknown or newly discovered numbers. This restriction is especially important when it comes to the Bangla language since current solutions are not sophisticated enough to distinguish between real and fake calls. As a result, people who speak Bangla are still quite susceptible to other types of telecommunications fraud, such as phishing, which involves tricking someone into giving up personal information. People who rely significantly on telecommunication services run the danger of suffering significant financial losses and serious security breaches due to the incapacity of present technologies to analyze and interpret voice scripts in Bangla. In addition to undermining consumer confidence, this inadequate fraud detection technology also

hinders the expansion and dependability of telecom services in Bangla-speaking areas.

1.4 Objective

The primary objective of this study is to create a sophisticated fraud detection system tailored to the Bangla language. The system will employ conversation transcript analysis to improve the precision and dependability of detecting fraudulent phone calls. The goal of this system is to overcome the shortcomings of conventional blacklist-based methods, which frequently fall short in identifying new or unidentified pattern phone numbers. The suggested method will evaluate voice scripts and use deep learning models and Natural Language Processing (NLP) techniques to successfully identify between legitimate and fraudulent calls. The research focuses on multiple important goals to accomplish this. First and foremost, find all necessary patterns of fraud phone call dataset. Second, the effectiveness of using and optimizing different deep learning models—ANN, RNN, LSTM, and BiLSTM, among others—for the analysis of recorded talks will be investigated. The ultimate goal is to construct a contemporary, trustworthy fraud detection system that maintains the integrity of telecommunication services and shields Bangla-speaking customers from money losses and security concerns by concentrating on the distinctive linguistic traits and nuances of the Bangla language.

1.5 Scope and Limitations

This research project extends its scope to address the threat of fraud attempts targeting users of mobile banking services such as Nagad and bKash and banking services as a whole, the telecommunication sector, e-commerce customers, and people who are vulnerable to financial loss to scam callers. With the increasing reliance on online banking, e-commerce shopping, and remote work opportunities, the susceptibility to fraudulent activities has elevated, necessitating a comprehensive approach to security. The project aims to deploy advanced voice analysis techniques to analyze different types of voice features within the Bengali language, thereby enabling the detection of fraud attempts that take advantage of unsuspecting users. By closely looking in on the unique challenges posed by this specific domain, the research explores different possible solutions to fortify the security measures of customer service systems, ensuring a strong defense against devious practices safeguarding the financial and well-being of users, and ensuring their security and privacy of information. This targeted scope builds a commitment to addressing the evolving scene of fraudulent activities, offering a specifically targeted solution to enhance the security state of customer safety and security throughout the domain.

There are several obstacles to overcome while putting into practice a task aimed at detecting fraudulent phone calls in Bangla. These include the absence of data and research in this area. The lack of a relevant dataset that is specifically focused on fraudulent phone calls in Bangla is one of the main problems. To obtain pertinent data in the absence of previous work in this field, it is imperative to directly connect with persons who have had such calls. Acquiring this data, nevertheless, is not without its difficulties because many people could be reluctant to divulge specifics of these chats out of discomfort or privacy concerns. The limited number of open-source Bangla ASR models exacerbates this problem. Word Error Rate (WER), which is often introduced by ASR models, becomes crucial in real-world applications requiring speech-based data. Creating a strong NLP model that can handle high WER and reliably identify fraudulent phone calls is crucial to reducing this. In order to guarantee proper categorization even in the event of probable errors generated during voice recognition, such a model would need to account for the subtleties of spoken Bangla, including changes in dialect and pronunciation. To put it simply, overcoming these obstacles calls for a multifaceted strategy that includes data-gathering techniques that are considerate of privacy issues, the development of ASR capabilities specifically tailored for Bangla, and the construction of robust NLP models that can function well even in not ideal speech-to-text scenarios. In order to successfully deploy solutions to combat fraudulent phone calls in the Bangla language environment, it will be imperative to address these obstacles.

1.6 Report Organization

The proposed report is structured with a comprehensive layout to guide readers through the research process, findings, and implications. Each chapter serves a specific purpose, contributing to the overall coherence and depth of the document.

Chapter 1: Introduction

The introductory chapter sets the stage by highlighting the growing threat of fraud in Bangladesh's telecommunications and mobile banking sectors. It outlines research objectives to develop advanced fraud detection systems for Bangla voice calls, posing specific questions on deep learning, fine-tuning models, predictive strategies, and NLP techniques. The chapter underscores the study's significance in enhancing telecommunication security and consumer protection, offering an overview of the report's structure.

Chapter 2: Literature Review

In the literature review, the goal is to present a thorough analysis of previous studies and published material that is pertinent to the identification of fraudulent phone calls. The purpose of this part is to explore the technological developments, methodology,

and theoretical foundations of deep learning models, fine-tuning strategies, and fraud detection. The background chapter aims to create a contextual framework for the current investigation by synthesizing this data. In addition, it seeks to pinpoint areas in which additional research is required to develop the field of fraud detection in telecommunications, with a special emphasis on enhancing precision and effectiveness through sophisticated AI techniques.

Chapter 3: Methodology

The study's detailed procedure is described in detail in the methodology chapter, which provides an overview of the model architecture, data gathering techniques, research design, and the reasons for their selection. To provide transparency and consistency in the analysis of fraud detection in telecommunications and mobile financial services, this section aims to shed light on the research methodology.

Chapter 4: Implementation

In this chapter, the Actual experiment is conducted. The models are used in different datasets. Describe every model implementation of fraud phone call detection. Their performance is analyzed with proper training and validation diagrams.

Chapter 5: Result And Analysis

This chapter presents the experimental results obtained from both a clean, manually supervised dataset and a transcribed audio dataset. It evaluates the performance of fine-tuning deep learning models using relevant metrics to assess their effectiveness in detecting fraud in telecommunications and mobile financial services.

Chapter 6: Impact on Society, Environment and Sustainability

This chapter examines the social implications of the fraudulent phone call detection system, outlining the ways in which people would be impacted on a personal and social level. This strategy protects the fraud detection system's long-term efficacy and dependability against changing scam techniques, fostering a safer online environment for all users.

Chapter 7: Conclusion and Future Work

The last chapter summarizes the contents of the full report. It presents a summary of the most important discoveries, restates the research's conclusions, and makes suggestions for real-world uses and additional research. Researchers who wish to build on the current work can find guidance in the discussion of the implications for future research. A logical flow of information is ensured by this organized layout, which leads the reader through the study process from the introduction to the suggestions and wider consequences.

1.7 Summary

In Bangladesh, MFS is growing rapidly. The extensive use of MFS has increased

fraudulent activities, including impersonation and account takeovers, leading to substantial financial losses and eroding confidence in these services. Conventional fraud detection techniques, which depend on CDRs, are inadequate in detecting complex frauds that include unknown phone numbers. This paper presents a sophisticated fraud detection system for individuals who speak Bangla. The system transcribes phone call conversations using ASR and then analyses the transcriptions using NLP techniques. To do this, advanced deep learning models like ANN, RNN, LSTM, and BiLSTM will analyze the transcriptions. Additionally, the research will focus on addressing the distinctive linguistic characteristics of the Bengali language. The objective is to develop a dependable system capable of discerning between authentic and deceitful calls, bolstering telecommunication services' security, and safeguarding Bangla-speaking consumers against monetary damages. In order to create reliable NLP models that can identify fraud even in noisy, real-world settings, the project also addresses the issues of data scarcity, privacy concerns, and the high Word Error Rate (WER) in ASR models.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview

This section examines the body of research on fraudulent phone call identification, emphasizing the dearth of studies specifically on Bangla. Two basic study methodologies are commonly used in this domain: the first involves analyzing Call Detail Records (CDR) and the other involves applying Natural Language Processing (NLP) techniques to analyze call content. The first method makes use of CDR datasets, which offer metadata including phone numbers, hours, locations, and call durations. To examine these databases and find trends and abnormalities that can point to fraud, researchers use machine learning algorithms. By using the structured nature of CDR data to identify unusual calling activities, this technique has been efficacious in identifying specific forms of fraud. The second strategy, which has gained popularity recently, is centered on the in-depth examination of call content. This is turning audio recordings into text and using natural language processing (NLP) to carefully examine the discussions for indications of fraud. Through the examination of language and conversational clues present in the transcripts, researchers can discern tiny symptoms of fraudulent behavior that may be overlooked through metadata analysis. With this approach, fraudulent behaviors can be understood more precisely and contextually.

2.2 Related Works

Jabbar and Suharjito[1] employed K-Means and DBSCAN algorithms to analyze Call Detail Records (CDRs) in their investigation of fraud detection in telecommunications employing machine learning. Their study sought to detect abnormalities that are suggestive of fraudulent activity, by evaluating the effectiveness of various algorithms using real fraud data. The findings demonstrated that the K-Means algorithm outperformed DBSCAN in accurately detecting fraudulent activity inside CDRs. This discovery highlights the effectiveness of the K-Means algorithm in grouping and identifying fraudulent activities, especially in situations that include extensive datasets commonly found in telecommunications operations. The authors emphasized the necessity of conducting more research on alternate clustering approaches and unsupervised learning methods, such as hierarchical grouping, in order to improve the effectiveness of fraud detection. They recognized the changing nature of fraud difficulties and the differences in how successful detection is for various telecommunications providers and fraud situations. They proposed continuous study to improve and broaden detection methods in this important field.

Xing et al. [2] present a sophisticated method for addressing fraudulent phone calls by

employing deep learning techniques. They tackle the constraints of conventional fraud detection algorithms that depend on manually designed characteristics, which might become outdated as fraud strategies advance. The authors propose a system that utilizes convolutional neural networks (CNNs), long short-term memory networks (LSTMs), and stacked denoising autoencoders (SDAEs) to automate the feature engineering process. These models are particularly trained to analyze CDRs and identify subtle trends in phone number features and call behaviors that are linked to fraudulent operations. Their experimental findings exhibit significant improvements in performance compared to traditional methods, obtaining an accuracy rate that is above 99%. Out of the models that were examined, the SDAE consistently performed better than the others, demonstrating higher stability in dealing with idea drift. Idea drift is an important factor where fraudulent behaviors change over time. The success of deep learning in adjusting to dynamic changes in fraudulent phone numbers and behaviors highlights its resilience, making it a strong solution for real-world applications in telecom security. The study by Xing et al. emphasizes the significant impact of automated feature learning on enhancing the accuracy, efficacy, and dependability of systems designed to identify fraudulent phone calls.

Zhao et al. The work of [3] for call content analysis with telecommunication fraud detection uses a model decision tree algorithm to build detection rules. They use the decision tree to ensemble select keywords; they contain fraud and normal data set-related coefficients in formulas. Traditional decision tree approaches emphasize information gain; in contrast, this research deals with the correlation value of all keywords for fraud detection. These correlation values will be summed over all keywords present in a text to create the detection rule, and if the sum exceeds a pre-set threshold, I classify the text as fraudulent; otherwise, it's normal. This emphasizes the use of decision trees in generating good rules to detect telecommunication fraud (using call content as input).

In their research, Moussavou and Park [4] propose a novel hybrid architecture that combines a 1-dimensional Convolutional Neural Network (1D CNN), Bidirectional Long Short-Term Memory (BiLSTM) and Hierarchical Attention Networks (HANs). This hybrid approach combines the power of CNNs and BiLSTMs to harvest context-aware features from word vectors ENCHMARKS The 1D CNN is successful at identifying local feature patterns in the input data, whereas BiLSTM well understands long-range dependencies along a sequence. Integrating with Hierarchical Attention Networks to improve the model by focusing on important features on the word and sentence levels. As a result of the evaluations performed on the real-world KorCCVi v2 dataset, this model achieved accuracy and F1 scores of 99.32 % and 99.31%, respectively, which are greater than any baseline models trained by the

authors. I will further show that the hybrid model outperforms many existing solutions, both those that are designed on structured grids and those tailored for unstructured grids. The results also demonstrate the comparable performance of the present approach to other approaches available in the literature, which verify its robustness and effectiveness.

In their study, Kumar et al. [5] concentrate on the detection of fraud calls by supervised learning algorithms. The research uses a dataset of 5,925 records. In this study, the performance of several machine learning algorithms (Random Decision Forest (RF), k-nearest Neighbours (k-NN), Support Vector Machine (SVM), and Gaussian Naive Bayes (GNB)) is being evaluated. Of all the k-NNs, SVM showed 99% accuracy. The paper discusses the need for early detection of scam calls to prevent a wide range of fraudulent scenarios and safeguard potential victims. The comparative analysis of various algorithms in this domain revealed that k-NN and SVM fare well, which strictly paves the way for a concrete predictive analytics-based fraud detection framework.

Naidu [6] collects telecommunication fraud records, including data from media and social platforms with the assistance of machine learning (ML) that filters and selects high-quality records. Use natural language processing (NLP) to extract features from the text, and define thresholds to flag like scammy call content. An Android application performs real-time fraud detection by dynamically evaluating the content of a call. The BERT model was the best performer of whatever models were tested from an acc and loss perspective. It gives an accuracy of 94.0% and a validation loss of 0.16. They use this model because it is the best at predicting whether a call is fraudulent or not.

A work by Neha[7] attempts to classify calls based on voice data transmitted between prospective victims and callers. The authors implemented two models and tested them with various machine-learning methods based on intent analysis of call transcripts. The model itself is built on Naive Bayes Algorithm and Convolution neural net, which is accurate up to 94.57% as well as 97.21%.

Natarajan et al. [8] present a novel deep-learning method for phone call (spam/ham) classification. They transcribe calls to text and then use embedded capabilities in a 2-Path Hybrid Model (2PHM). It combines a Long LSTM with a Sentence Similarity Network. The 2PHM, which utilizes the Universal Sentence Encoder embedding technique, transforms call transcripts into latent feature vectors in a high-dimensional embed space. On the other hand, their model is more accurate than well-known Machine Learning models like Naïve Bayes, K-Nearest Neighbors, Random Forests, and Support Vector Machines with a classification accuracy of 93% and a runtime of 3 seconds.

With a particular focus on call center applications, Ezzat et al.[9] investigates the use of text mining techniques on transcribed audio recordings to discern the moods of speakers. The suggested approach entails employing speech recognition technology to transcribe audio, converting the text to a "bag-of-words" representation, and using a variety of feature selection techniques to find pertinent keywords. Next, the study extracts distinguishing keywords, clusters comparable calls, and classifies calls based on sentiment using supervised and unsupervised learning approaches such as Support Vector Machines (SVMs), Naïve Bayesian classifiers, decision trees, and K-nearest neighbor classifiers. The strategy looks for terms linked to poor customer service and works to enhance automated systems that rate customer support conversations. Between all the models SVM gives maximum accuracy of 94%

In another work, deep-learning neural networks were used to detect spoofed call center conversations [10]. Their most successful results use deep CNNs with word embedding vectors as inputs. The former allows 62% of fake calls to be identified automatically with 43% percent accuracy.

Mawgoud et al.[11] suggest to use of artificial neural networks (ANN) for Wangiri fraud detection of telecom datasets. Wangiri fraud which makes missed calls to lure an expensive calling back is one of the threatening trends among has increased danger in the telecom industry. For the dataset, authors took 1000 samples in total including 700 as normal cases and 300 as fraudulent cases. The dates and times in the data set are converted to facet variables. A method tested a Panama-based dataset using an ANN model with MLP and achieved 90.6% success for the telecom dataset and 85% on the test data set. The research reveals the power of neural networks for recognizing fraudulent patterns, even without any human interaction, and suggests using them to improve telecom fraud detection.

Hong, B. et al[12] used Google API to transcribe the call contents and then apply word embedding (and other NLP tasks) creating a more reliable classifier. The Long Short Term Memory (LSTM) model was developed, and the highest accuracy of 85.61% was achieved in detecting scam calls.

Elizalde & Emmanouilidou[13] proposed shallow feature-based detection using call detail records while their methods examine acoustic voicemails for significant performance improvement over traditional anomaly-based phone call spam detection. Their method can determine whether or not a call is from a human with 93% accuracy, subsequently filtering out spam calls as long as they are coming from humans with between 75-83% accuracy. This privacy-preserving approach, which does not analyze the content of spoken words, demonstrates that voiced and unvoiced audio content alone contains enough information to distinguish between call types. This research leaves room for the broader use of this audio-based approach, to

complement existing call behavior analytics, to minimize unwanted and dishonest calls.

2.3 Comparison between existing works

Table 2.3.1: Comparative analysis of some work for Fraudulent Phone Call Detection

SL No	Author Name	Used Algorithm	Best Accuracy with Algorithm
1.	Jabbar & Suharjito [1]	K-Means, DBSCAN	K-Means = 97%
2.	Xing et al.[2]	CNN, LSTM, SDAE	CNN = 99%
3.	Zhao et al.[3]	Decision Tree	Decision tree = 98%
4.	Moussavou & Park [4]	ANN, BiLSTM,HAN	Hybrid(ANN, BiLSTM,HAN) = 99.32
5.	Kumar et al.[5]	Random Decision Forest (RF), k-nearest Neighbours (k-NN), Support Vector Machine (SVM), and Gaussian Naive Bayes (GNB).	SVM = 99%
6.	Naidu[6]	LSTM, BiLSTM, BERT	BERT = 94%
7.	Kale et al. [7]	Naive Bayes, CNN	CNN = 97%
8.	Natarajan et al. [8]	Naive Bayes, KNN, LSTM, Support Vector Machine	Hybrid Model = 93%
9.	Ezzat et al.[9]	Naive Bayes, Decision Tree,, SVM	SVM=94%
10.	Özlan et al. [10]	Naive Bayes, Decision Tree, CNN	CNN = 43%
11.	Mawgoud et al. [11]	Multilayer perceptron (MLP) ANN	Multilayer perceptron (MLP) ANN = 85%

12.	Hong et al. [12]	SVM,CNN,LSTM	LSTM = 85.61%
13.	Elizaldz et al. [13]	SVM, L-SVM, CNN	CNN=93%

2.4 Open Issues

Even with these developments, research on the topic of detecting fraudulent phone calls in the Bangla language is still lacking significantly. The distinct linguistic features of Bangla provide certain difficulties that have not been adequately addressed in previous research. There is also an absence of performance analysis in real-world scenarios. By using cutting-edge speech analysis methods designed especially for Bangla, this paper seeks to close this gap and improve fraud detection's precision and efficacy in this linguistic setting. Moreover, the influence of imperfect real-time transcriptions on the performance of NLP models has not been sufficiently investigated in prior research.

2.5 Summary

Two main approaches are used in research on fraudulent phone call identification: the analysis of Call Detail Records (CDR) and the use of Natural Language Processing (NLP) techniques to call content. CDR analysis uses machine learning to identify fraud tendencies by utilizing metadata such as phone numbers and call durations. NLP techniques use linguistic signals to translate audio recordings into text in order to detect fraud. Although these techniques have proven successful in identifying different kinds of fraud, there is a huge research vacuum concerning Bangla. The distinct language characteristics of Bangla provide particular difficulties that have not been sufficiently addressed. Furthermore, more research is necessary to determine how imperfect real-time transcriptions affect the performance of NLP models.

CHAPTER 3
METHODOLOGY/ REQUIREMENT ANALYSIS & DESIGN SPECIFICATION

3.1 Overview

This section will present my suggested strategy for identifying phoney calls, along with in-depth information on the methods used for data collection, data augmentation, and data preprocessing. I'll list the gear and software I need for my job to guarantee scalability and the best performance. This part also discusses financial analysis and project management techniques, giving readers a comprehensive understanding of the project's resources, scope, and financial concerns. By applying a rigorous methodology to both technical implementation and project management, my goal is to create a reliable and affordable system for detecting phoney calls.

3.2 Proposed Methodology

The primary objective of my suggested methodology is to identify fraudulent telephone conversations by translating spoken language into written text and subsequently analyzing the transcribed material using NLP. The process of transcription conversion entails the utilization of ASR technology. Subsequently, I employ natural language processing (NLP) methods and deep learning algorithms to identify deceitful phone conversations by analyzing the transcribed text. The execution of my technique is carried out in two distinct stages. I first trained my model on a clean text dataset that was manually written after conducting interviews with people who faced fraudulent phone calls. This dataset comprises accurately labeled text samples that are devoid of any environmental interference or transcription mistakes. The primary purpose of this first training is to enable the model to acquire knowledge about the fundamental patterns and characteristics that are linked to both fraudulent and non-fraudulent calls. However, when the model is used in real-life situations, the transcription data will inevitably include noise caused by variables like a loud environment and the intrinsic Word Error Rate (WER) of the ASR model. In order to tackle this issue, I optimize my pre-existing NLP models by utilizing the transcribed text produced by my ASR system. This step allows my model to account for variations and inaccuracies in real transcriptions. I will try to assess different kinds of deep learning models on the transcribed dataset, which I have fine-tuned to ground truth annotations. This experimental contrast helps us to better understand how fine-tuning will affect the accuracy and robustness of a model with noisy real-world transcriptions. Their rigorous approach means that the models are able to deal with practical scenarios well, even if the data quality is not great.

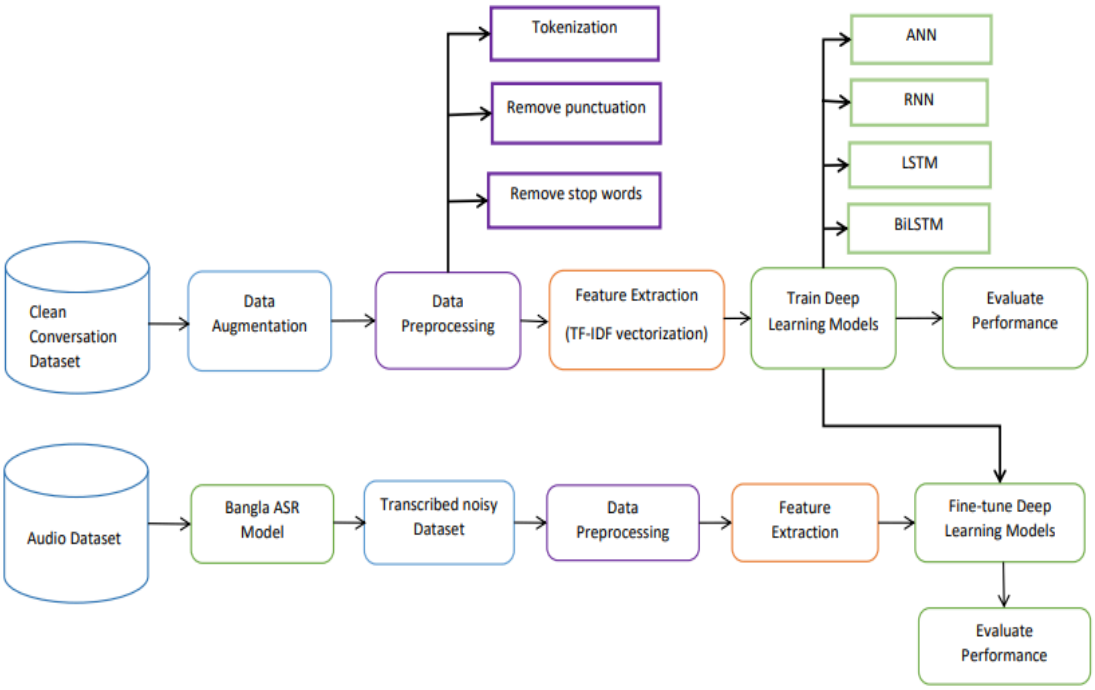


Figure 3.2.1: Proposed Methodology

3.2.1 Data Collection Procedure

The success of fraud detection depends to a large extent on the quality of the underlying data, so careful dataset preparation is necessary. The dataset includes 120 phone call scripts, distributed evenly between fraud and non-fraud ones. This data was gathered by analyzing valid newspaper sources and interviews with people who had received scam phone calls. These scripts were manually transcribed, ensuring accurate word representation. This thorough procedure ensures both the correctness and the appropriateness of my dataset for fraud detection methods with positive predictive value. It is important to note that my analysis focuses exclusively on the conversations initiated by the scammers. Therefore, I have analyzed only conversations or callers and by using different models I will identify scammers. This systematic approach ensures that the next analyses will be concentrated on the relevant discourse necessary to understand patterns and methods of fraudulent communication. A sample of the dataset is illustrated in Figure 3.2.1.1 To evaluate the performance of NLP models on transcriptions produced by the ASR model, I have gathered a total of 480 recordings of different individuals engaging in these conversation scripts from the callers' perspective. A sample of the recording transcription dataset is illustrated in Figure 3.2.1.2 This comprehensive dataset serves as the foundation for my analysis and model assessment.

Table 3.2.1.1: Table for some Clean Conversations from my collected Dataset

Callers Conversation	Type
আমি সোনালি ব্যাংক এর হেড অফিস থেকে বলছি। আমরা প্রতি তিন মাস পর পর নিবন্ধিত একাউন্ট গুলোর উপর রেফাল ড্র করি। এই ড্র তে আপনি অষ্টম পুরষ্কার হিসেবে ১৫০০০০ টাকা পেয়েছেন। না স্যার। টাকা আপনার কালকে অনুষ্ঠানের মাধ্যমে সবাইকে বুঝিয়ে দেওয়া হবে। আপনি চাইলে আমি ম্যানেজারের সাথে কথা বলে টাকাটা আপনাকে বিকাশ করতে পারি। বিভিন্ন হাত হয়ে টাকা আনতে হয় আমাদের। অনেক জায়গার বখশিস দিতে হয়।দেখুন, আমি কোনো টাকা রাখবো না বাট অনেক জন কে দিতে হবে। এখন আমি নিজের পকেট থেকে তো দিতে পারবো না। তো আপনি যদি চান তাহলে একটা বিকাশ একাউন্ট এ ৫০০০ টাকা সেন্ড মানি করুন। তারপর আমি টাকা টি তুলে আপনার একাউন্ট এ বিকাশ করে দিব।	fraud
হ্যালো, আসসালামুঅালাইকুম, আমি টেন মিনিট স্কুল থেকে [নাম] বলছি। আমরা এইচএসসি শিক্ষার্থীদের জন্য ইংরেজি কোর্স চালু করেছি। আপনি কি কোর্সটি করতে ইচ্ছুক?? আপনার পরিচিত কোনো ছোট ভাইবোন থাকলে কোর্সটি রিকমেন্ড করতে পারেন। ঠিকাছে স্যার, আমি মেসেজ করে দিচ্ছি। আমাকে সময় দেওয়ার জন্য আপনাকে ধন্যবাদ।	non-fraud

Table 3.2.1.2: Table for some transcribed Conversations from my collected Dataset

Transcribed Conversation	Type
আমি সোনালী ব্যাংকের হেড অপিস থেকে বলছি, আমরা প্রতিন তিন মাস পরপর উপর রেফাল রগ গড়ি, এই ড্রতে আপনি চলকে পাশকলনের করারণ কর ◆অষ্টম পুরস্কার হিসাবে, এক লাখ তাকা পেয়েছেন, না সর টাকাটা আপনার কালকে মাধ্যমে সবাইকে বচিয়ে দেয়া হবে।আপনি চাইলে আমি সাথে কথাকাটা আপনাকে বিক্রিকে হাত হয়ে তাকা আন্ত হয়ে আমাদের, অনেক জায়গের বক্সেই দিতে হয়ে, দেখন আমি কোন তাকার রাত বোন কোন, কিন্তু অনেক জনকে দিত্ব হবে। এখন আমি নিজের পকেটে থেকার তাকার ক্যবার ক্যতাকা দিতে পার মনা, তো আপরিযোদি চান, তাহলে একটি বিকাশ একাউন্টে পাঁচ হাজার তাকা শ্যান মানি করেন, তারপর আমির তাকাটি তুলে আপনার ব্যাঙ্ক একাউন্টে বিকাশ করে দিগব।	fraud
হ্লো, আসলাম ওয়ালেকম, আমি টেন মেয়েট স্কুল থেকে বলছি, আমরা এইটিসি জন্য ইংরেজি কোষ চালু করেছি, আপনিকি কোষ তৈই করতে ইচ্ছুক, আপনিকি করত্য কোন ছোট ভাইয়ে ইবন থাকলে কোস্ত্রি কমেন্ড করতে পারেন।টিক আচ্ছেস করে লিচছি, আমাকে সময় দয়হের জন্য আপনায়কে	non-fraud

3.2.2 Data Augmentation

Given the limitations of my initial dataset, To reduce the chance of overfitting, I have used data augmentation techniques on my cleaned dataset. Specifically, I have utilized similar alternative word augmentation techniques, which generate new scripts by substituting words with semantically similar alternatives. This technique leverages pre-trained word embedding models (BengaliWord2Vec) to create a relationship for words in some semantic manner. I have used word synonyms from words having high

similarity scores to synthetically grow my dataset without altering the core sense of the conversations. Combining two types of network inputs, allows the model to pick up more durable representations and reduces the risk of overfitting to only one category of original data. Top 3 Similar alternative words for 'টাকা' are 'রুপি', 'পয়সা', 'টাকার'. I have implemented it using a BNLP (Bangla Natural Language Processing) library. After Augmentation, my data points increased to 1920 where 960 is fraud script and 960 is non-fraud script.

Table 3.2.2.1: Comparison between Real & Augmented Conversation from my Dataset

Real Conversation	Augmented Conversation
আমি সোনালি ব্যাংক এর হেড অফিস থেকে বলছি। আমরা প্রতি তিন মাস পর পর নিবন্ধিত একাউন্ট গুলোর উপর রেফাল ড্র করি। এই ড্র তে আপনি অষ্টম পুরষ্কার হিসেবে ১৫০০০০ টাকা পেয়েছেন। না স্যার। টাকা আপনার কালকে অনুষ্ঠানের মাধ্যমে সবাইকে বুঝিয়ে দেওয়া হবে। আপনি চাইলে আমি ম্যানেজারের সাথে কথা বলে টাকাটা আপনাকে বিকাশ করতে পারি। বিভিন্ন হাত হয়ে টাকা আনতে হয় আমাদের। অনেক জায়গার বখশিস দিতে হয়।দেখুন, আমি কোনো টাকা রাখবো না বাট অনেক জন কে দিতে হবে। এখন আমি নিজের পকেট থেকে তো দিতে পারবো না। তো আপনি যদি চান তাহলে একটা বিকাশ একাউন্ট এ ৫০০০ টাকা সেন্ড মানি করুন। তারপর আমি টাকা টি তুলে আপনার একাউন্ট এ বিকাশ করে দিব।	আমরা গোলাপি ব্যাংকের এটির ডাইরেক্টর কার্যালয় থেকেও বলছি। আমি প্রতিও চার সপ্তাহ পরে পরে তালিকাভুক্ত অ্যাকাউন্ট সমূহের ওপর রেফাল ১-১ করি। এসব ১-১ য় আমি ষষ্ঠ পুরস্কার হিসেবেও ৬,০০০ রুপি পেয়েছেন।না স্যার। রুপি আমার গঠনই উৎসবের জন্য সকলকে বুঝিয়ে দেওয়া হবে। আমি চাইলেও আমরা ম্যানেজারের সঙ্গে কথাও বলেও ফিরিয়েও আমাকে প্রসার করতেও পারি। নানা আঁকড়ে হয়ে রুপি আনেন হয় আমাদের। বহু জায়গার ম্যানেজারের দিতেও হয়।দেখুন, আমরা কোন রুপি রাখবো নি লাকি বহু জনকে -কে দিতেও হবে। আজ আমরা তার বাক্স থেকেও তুমি দিতেও পারব নি তুমি আমি তাহলে চেয়েছিলেন তাহলেও একটি প্রসার অ্যাকাউন্ট এই ৪০০০ রুপি বার্বেট লন্ডারিং করুন। এরপর আমরা রুপি টির এনে আমার অ্যাকাউন্ট এই প্রসার করবে দিব।
হ্যালো, আসসালামুঅালাইকুম, আমি টেন মিনিট স্কুল থেকে [নাম] বলছি। আমরা এইচএসসি শিক্ষার্থীদের জন্য ইংরেজি কোর্স চালু করেছি। আপনি কি কোর্সটি করতে ইচ্ছুক?? আপনার পরিচিত কোনো ছোট ভাইবোন থাকলে কোর্সটি রিকমেন্ড করতে পারেন। ঠিকাছে স্যার, আমি মেসেজ করে দিচ্ছি। আমাকে সময় দেওয়ার জন্য আপনাকে ধন্যবাদ।	হ্যালো, আসসালামুঅালাইকুম, আমি টেন ঘন্টা কলেজ থেকে বলছি। আমরা এসএসসি ছাত্রদের জন্য ইংরাজি ডিপ্লোমা চালু করেছি। আপনি কি পাঠক্রম করতে ইচ্ছুক?? আপনার খ্যাত কোনো ছোটো ভাই-বোন থাকায় পাঠক্রম রিকমেন্ড করতে পারেন। ঠিকাছে স্যার, আমি পাসওয়ার্ড করে দিচ্ছি। আমাকে সময় দেওয়ার জন্য আপনাকে ধন্যবাদ।

3.2.3 Data Preprocessing

For advancing further into machine learning and deep learning, which are used in NLP, proper data preprocessing is very important. It requires a methodical process to clean the data and prepare it for analysis. I have hidden some of the personal information of the interviewees in my dataset with special symbols to preserve privacy during preprocessing. The names of the institutions were also removed to

prevent any potential biases. Apart from the concerns of privacy, the corpus has been enhanced by carefully eliminating punctuation and stop words, guaranteeing that the final dataset is well-suited for natural language processing applications. Most of the preprocessing tasks use a Python library called "BNLP," which allows you to systematically clean the data and make it into a format in which you want to train your model, while also helping to improve the performance of the model and providing accurate analysis in NLP applications. Table 3.2.3.1 shows the comparison between original and preprocessed clean corpus.

Table 3.2.3.1: Comparison between Original & Clean Corpus from my dataset

Original Dataset	**Clean Corpus**
আমি সোনালি ব্যাংক এর হেড অফিস থেকে বলছি। আমরা প্রতি তিন মাস পর পর নিবন্ধিত একাউন্ট গুলোর উপর রেফাল ড্র করি। এই ড্র তে আপনি অষ্টম পুরষ্কার হিসেবে ১৫০০০০ টাকা পেয়েছেন। না স্যার। টাকা আপনার কালকে অনুষ্ঠানের মাধ্যমে সবাইকে বুঝিয়ে দেওয়া হবে। আপনি চাইলে আমি ম্যানেজারের সাথে কথা বলে টাকাটা আপনাকে বিকাশ করতে পারি। বিভিন্ন হাত হয়ে টাকা আনতে হয় আমাদের। অনেক জায়গার বখশিস দিতে হয়।দেখুন, আমি কোনো টাকা রাখবো না বাট অনেক জন কে দিতে হবে। এখন আমি নিজের পকেট থেকে তো দিতে পারবো না। তো আপনি যদি চান তাহলে একটা বিকাশ একাউন্ট এ ৫০০০ টাকা সেন্ড মানি করুন। তারপর আমি টাকা টি তুলে আপনার একাউন্ট এ বিকাশ করে দিব।	ব্যাংক হেড অফিস তিন মাস নিবন্ধিত একাউন্ট গুলোর রেফাল ড্র ড্র তে অষ্টম পুরষ্কার হিসেবে টাকা পেয়েছেন না টাকা কালকে অনুষ্ঠানের সবাইকে বুঝিয়ে চাইলে ম্যানেজারের কথা টাকাটা হাত টাকা আনতে জায়গার বখশিস টাকা রাখবো বাট পকেট পারবো একটা একাউন্ট টাকা সেন্ড মানি টাকা একাউন্ট
হ্যালো, আসসালামুঅালাইকুম, আমি টেন মিনিট স্কুল থেকে [নাম] বলছি। আমরা এইচএসসি শিক্ষার্থীদের জন্য ইংরেজি কোর্স চালু করেছি। আপনি কি কোর্সটি করতে ইচ্ছুক?? আপনার পরিচিত কোনো ছোট ভাইবোন থাকলে কোর্সটি রিকমেন্ড করতে পারেন। ঠিকাছে স্যার, আমি মেসেজ করে দিচ্ছি। আমাকে সময় দেওয়ার জন্য আপনাকে ধন্যবাদ। আপনার পরিচিত কোনো ছোট ভাইবোন থাকলে কোর্সটি রিকমেন্ড করতে পারেন। ঠিকাছে স্যার, আমি মেসেজ করে দিচ্ছি। আমাকে সময় দেওয়ার জন্য আপনাকে ধন্যবাদ।	মিনিট স্কুল এইচএসসি শিক্ষার্থীদের ইংরেজি কোর্স কোর্সটি পরিচিত ছোট ভাইবোন থাকলে কোর্সটি রিকমেন্ড ঠিকাছে মেসেজ সময় দেওয়ার পরিচিত ছোট ভাইবোন থাকলে কোর্সটি রিকমেন্ড ঠিকাছে মেসেজ সময় দেওয়ার

3.2.4 Feature Extraction

The primary features that other researchers retrieved from their approaches of

detecting telecom fraud include calling numbers, calling hours, calling types, etc. I will identify fraudulent calls using call content identification in my study. So I have to extract features out of more and more complex data, a problem of extraction method. TF-IDF vectorizer was used for the feature extraction. TF-IDF name stands for term frequency-inverse document in its entire form.

$$TF(x) = \left(\frac{Amount\ of\ times\ term\ x\ appears\ in\ a\ specific\ document}{Total\ amount\ of\ terms\ in\ that\ document} \right) \quad \ldots\ldots\ldots (1)$$

$$IDF(x) = \log\left(\frac{Total\ Amount\ of\ Documents}{Amount\ of\ Documents\ with\ term\ x\ in\ it} \right) \quad \ldots\ldots\ldots (2)$$

TF-IDF = TF * IDF

The weighted values that TF-IDF gives us indicate the relative relevance of each attribute. The data is weighted and ready for use in the classification algorithm after all of these steps are completed. The stopwords from the dataset are eliminated throughout the preparation stages. The TF-IDF vectorizer is used to transform the remaining words into a numerical vector.

3.2.5 Model Selection & Design

Here, I have attempted to implement four advanced deep-learning models. The models that I have implemented in my work are:

- ANN
- RNN
- LSTM
- BiLSTM

Artificial Neural Networks (ANNs): ANNs are machine learning models that draw inspiration from the structure and functionality of the human brain. Neural networks are composed of linked nodes, known as "neurons," that are structured to analyze and acquire knowledge from inputs via weighted connections. Artificial neural networks (ANNs) are designed for specific purposes, such as pattern recognition or data categorization. These networks modify their connections through a training process to enhance their ability to accurately anticipate events. Backpropagation is a learning method that enables artificial neural networks (ANNs) to adjust and enhance their performance using known instances. This ability makes ANNs very proficient in dealing with intricate and non-linear correlations in data. Artificial neural networks (ANNs) are used in a wide range of fields, such as image and speech recognition, natural language processing, and predictive analytics. ANNs are capable of making educated judgments and predictions by drawing on their training data and using generalization techniques. Research is still being conducted to improve ANN structures, algorithms, and applications, which will increase their contribution to the development of machine learning and artificial intelligence technologies [16].

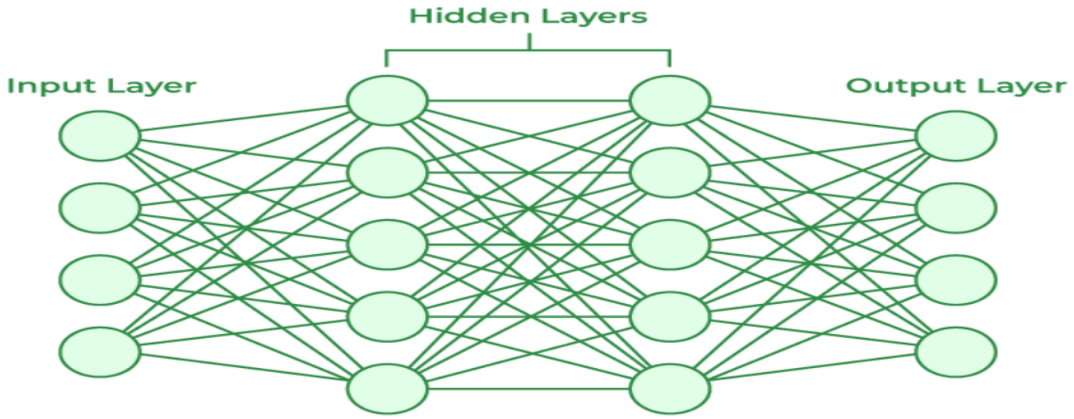


Figure 3.2.5.1: Architecture of ANN [16]

Recurrent Neural Network (RNN): RNN is a specialized neural network that handles sequential data by retaining a recollection of previous inputs in its hidden state. RNNs use this memory to evaluate context and dependencies across time, unlike regular neural networks. Predicting the next word in a phrase or analyzing time series data requires knowing temporal connections. It can catch patterns and make predictions based on sequential data because its hidden state stores information from prior inputs. RNNs are simpler than feedforward neural networks since they share parameters across time steps. Parameter sharing speeds up sequence learning and facilitates sequential processing applications. RNNs might struggle to learn long-term dependencies due to disappearing or ballooning gradients during training. LSTM and GRU RNNs improve memory retention and gradient flow control to overcome these issues. RNN designs and training methods are being improved to improve natural language processing, time series analysis, and other applications [17].

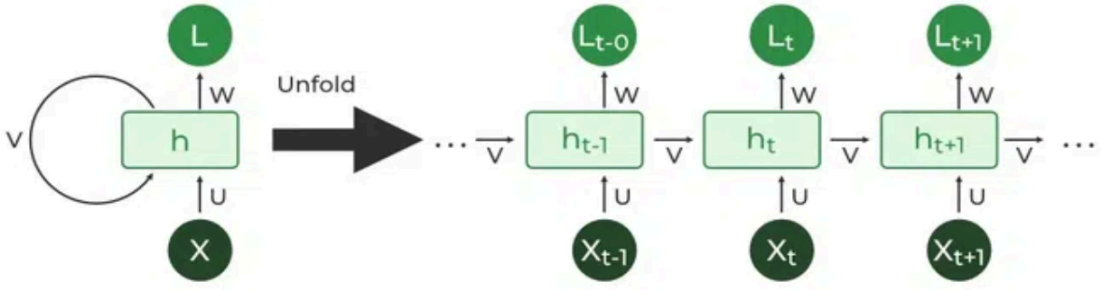


Figure 3.2.5.2: Architecture of RNN [17]

Long Short-Term Memory (LSTM): LSTM networks are a type of recurrent neural network that is specifically developed to address the difficulties associated with collecting long-term relationships in sequential data. Traditional RNNs have trouble keeping useful information over time because of problems like disappearing

gradients, which make it hard for them to learn from data from a long time ago.LSTMs overcome these restrictions by offering a distinctive design that includes gated cells. Cells are made up of input, forget, and output gates that regulate the information flow in the network. The input gate regulates the incorporation of fresh data into the memory; the forget gate oversees the selection of information to be discarded; and the output gate decides the ultimate output by considering the updated memory state. This organized method allows LSTMs to successfully capture and utilize long-term dependencies, making them excellent for applications like voice recognition, language modeling, and time series prediction that require understanding context over lengthy durations [18].

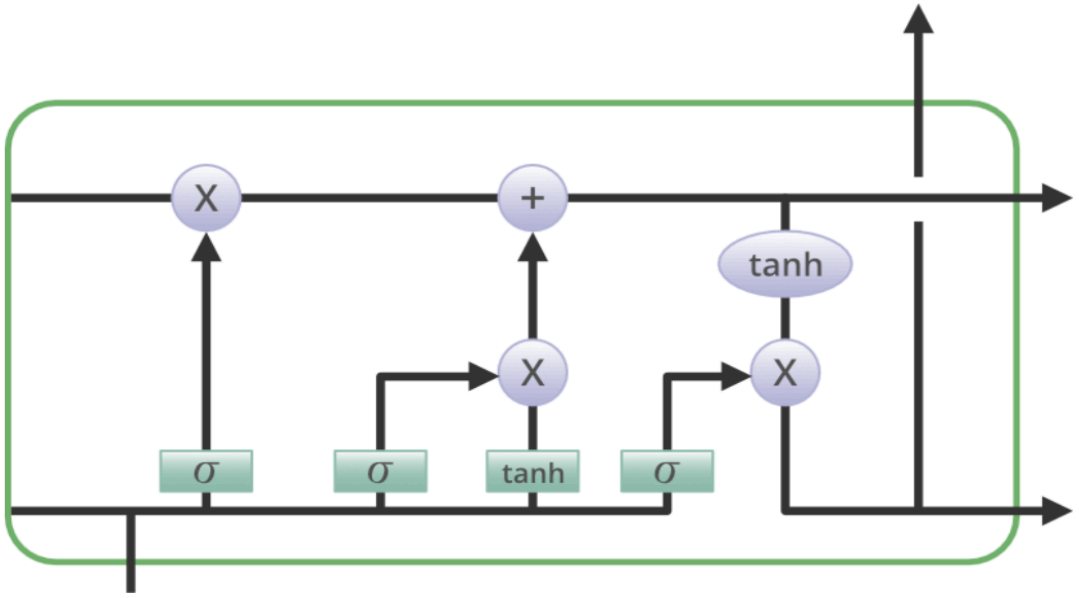


Figure 3.2.5.3: Architecture of LSTM [18]

Bidirectional LSTM (BiLSTM): BiLSTM improves upon regular LSTM networks by concurrently processing input sequences in both forward and backward orientations. The use of this dual technique enables BiLSTM to effectively capture contextual dependencies from both preceding and succeeding stages of the sequence. This makes it especially advantageous for tasks that require a comprehensive comprehension of the links between words in both directions.BiLSTM is particularly effective in natural language processing (NLP) for applications such as sentiment analysis, text categorization, and machine translation. For example, in sentiment analysis, the interpretation of a word might differ depending on the context in which it is used. The BiLSTM's capacity to take into account both preceding and subsequent words enhances its comprehension of the sentiment conveyed in a phrase. The architecture of BiLSTM consists of two LSTM layers operating autonomously, with one processing the sequence in its original order and the other processing it in reverse order. After that, the outcomes from both channels are combined to give a fuller

picture of the input sequence. This makes it easier and more accurate to see how complex connections exist within sequential data [19].

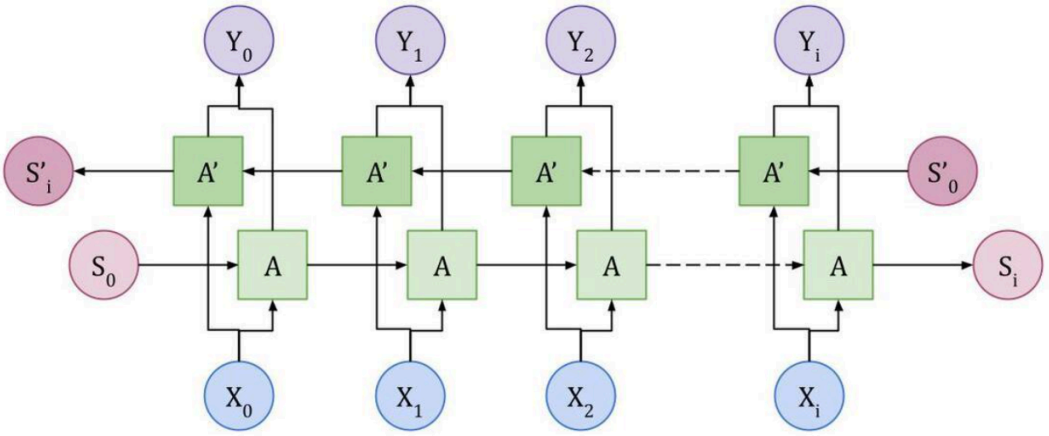


Figure 3.2.5.4: Architecture of BiLSTM [19]

Fine Tuning: Fine-tuning is the process of using the knowledge that an existing model has already acquired as the foundation for learning new tasks. It is a subset of the more general transfer learning technique**.** Fine-tuning refers to taking a pre-trained model and further training it on a new dataset. Using the base model's previous knowledge as a starting point, fine-tuning tailors the model by training it on a smaller, task-specific dataset. To avoid big changes, the learning rate for the first few layers is set to a low number. In contrast, the later layers use a higher learning rate in order to adjust to the updated dataset[24].

3.3 Hardware & Software Requirements

Hardware Requirements:

- **High-performance computing resources:** Due to the involvement of deep learning models and machine learning models, the usage of fast, high-memory, powerful GPUs was necessary.
- **Graphics Processing Unit (GPU):** A GPU is more effective when it comes to training more machine learning models compared to a CPU. This is why powerful GPUs are heavily used in the fields of AI and machine learning.

Software Requirements:

- **Bangla ASR Model:** I have used Hugging Face's Bangla ASR model for the transcribing process. This model is based on the Whisper architecture and has been specifically adjusted for the Bangla language. Here are some essential model details: The model was developed using 400 hours of audio data from the Bangla Mozilla Common Voice Dataset, which included 7,000 validation

samples and 40,000 training samples. A small variation of Whisper with 244 million parameters. After 12,000 steps of training, the model's Word Error Rate (WER) was 4.58%[22].

- **Deep learning models:** Since, the processing of natural language is involved in this study, deep learning models were used as the primary models.
- **Python programming language:** Python has a great ecosystem for working with AI and machine learning. The collection of rich libraries and huge communities of machine learning experts online make it easier and faster to work with machine learning.
- **Web Application:** For my work, I have built a web application that can be used to detect fraud and non-fraud conversation call conversations from call recordings. In the front end, I have used HTML, CSS, and Javascript. In the backend, I have used the FLASK framework of Python.

3.4 Project Management & Financial Analysis

3.4.1 Project Management

- **Project Objectives**
  A. To develop a model that can detect fraud from Bangla voice call
  B. develop a pipeline between the ASR model and the NLP model
  C. Make a system of combining the ASR model and the NLP model to detect fraud phone call
- **Project Timeline**
  - Phase 1: Project Planning and Research (10 weeks)
  - Phase 2: Raw Dataset Collection (12 weeks)
  - Phase 3: Preprocessing (8 weeks)
  - Phase 4: Model Implementation (8 weeks)
  - Phase 5: Result Analysis and Improvement (6 weeks)
  - Phase 6: Simple Software Design (4 weeks)
- **Resource Planning**
  A. Equipment and Tools
    - Development and Testing Servers
    - High-performance Computers for Development Team Design Software (e.g. Google Colab, Google Drive)
    - Version Control System (e.g., Git)
    - Plagiarism Checking Tools (e.g., Turnitin)
    - Highly capable computer with a good amount of RAM and GPU

  B. Software and technologies

- Front-End Technologies: HTML, CSS, JavaScript
- Back-End Technologies: Python FLASK
- Libraries: BNLP Library, Tensorflow, numpy, pandas
- Models: Bangla ASR model for speech-to-text transcription

C. Data and Content

- Voice call records collected from people
- Phone call scripts collected from people

D. Training and Skill Development

- Fundamental knowledge of deep learning
- Fundamental Knowledge of Natural language processing
- Understanding of transformer models
- Ability to use several tools, like Google API and Open AI products
- Understanding of the potential use cases of the final product

3.4.2 Financial Analysis

The project has several parts that require monetary supplements, and there are several phases to its life cycle where funding is necessary. In the research and development process, there is the necessity of voice transcription, There are tools available for the task for a price. Training the models requires huge amounts of processing power, and the hardware also requires funding.

Table 3.4.2.1: Estimated cost for this work

SN	Components	Estimated Cost (BDT)
01.	GPU for training	185,000
02.	Software and tools (Google API)	1000-2000
03.	Data Collection and Processing	40,000
04.	Contingency	1000-2000
	Total Estimated Cost	227,000-229,000

3.5 Summary

This study seeks to detect fraudulent phone conversations by transcribing spoken language into text using ASR and analyzing the transcriptions using Natural Language Processing (NLP) techniques. Hand-transcribed interviews with scam call recipients were the first step in creating a clean text dataset to ensure accuracy. Subsequently, this dataset was enlarged using data augmentation techniques, which

involved developing scripts that were semantically comparable to mitigate the risk of overfitting. The preprocessing step entailed anonymizing sensitive data and eliminating redundant components using the BNLP package. Features were retrieved with a TF-IDF vectorizer, emphasizing the significance of distinct words. ANN, RNN, LSTM, and BiLSTM are the four deep-learning models that were used to handle the sequential data and identify long-term relationships. High-performance GPUs were used together with a range of software tools, such as the Bangla ASR model. The project, which had a 48-week duration and six phases, required careful resource planning and had an anticipated cost of between BDT 227,000 and 229,000. This comprehensive approach guarantees that the models are resilient and efficient in practical situations.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

This section describes how my system was put into practice. The application of models on two different datasets is covered in subsection 4.2, along with an explanation of how my system handles them. It provides an overview of the process of choosing models and integrating them, including fine-tuning. The model implementation specifics are shown graphically in sub-section 4.3, which also highlights the model's performance on training and validation datasets.

4.2 Train Model

In this sub-section, I will use the conversational dataset to train the deep learning model. My two datasets are a transcription of audio files produced by Bangla ASR and a manually created call script. In my first experiment, I will use the manually written call script dataset to train my deep learning models. Next, I will use the transcription dataset to fine-tune my models.

4.2.1 Train models using a clean conversational dataset

The experimental setup used to train deep learning models on a conversational dataset gathered through one-on-one interviews is described in depth in this section. Three subsets of the dataset were created: test, validation, and training data. In particular, training accounted for 70% of the data, with the remaining 30% being split equally between test and validation sets. In the phases of testing and model validation, this guarantees balanced evaluation metrics.

I have trained four deep learning models: ANN, RNN, LSTM, and BiLSTM. Each model was trained using the training subset, with performance monitoring on the validation subset to optimize hyperparameters and prevent overfitting. The final evaluation was conducted using the test subset, providing an objective measure of each model's generalization capability.

Artificial Neural Network (ANN)

An Artificial Neural Network (ANN) trained on text scripts is used to classify fraudulent phone calls. The ANN is trained to identify patterns in the text that point to fraudulent activities. The ANN is made up of several layers of interconnected nodes, each of which processes the incoming data to extract progressively more sophisticated properties. The input layer uses the Relu activation function and L2 regularization to prevent overfitting. The network interprets a phone call script layer by layer as it is fed into the trained artificial neural network. To generate an output, each node in the network gives its input weight and runs it through an activation function. The hidden layers are where this process continues, detecting and amplifying patterns linked to

fraudulent activity. The last layer of the ANN's classification output typically provides the likelihood that the call is genuine or fake.

Figure 4.2.1.1: Working Flow of ANN for my work

Recurrent Neural Network (RNN)

This RNN is made up of multiple layers of connected nodes that process incoming input incrementally to extract increasingly complex features. Initially, the data is reshaped to meet the input specifications of the RNN, making sure it is appropriately organized for sequential processing. The model starts with an RNN layer that uses L2 to prevent overfitting. Another RNN layer is then added, and both are set up to return the entire sequence of data that has been processed. An extra RNN layer comes next, which carries on the job of finding and enhancing patterns connected to fraudulent activity. In order to further reduce overfitting, a dropout layer is included, which randomly removes nodes from the training set. The network then has thick layers with ReLU activation functions to further enhance the patterns that have been learned. The ultimate output layer produces a classification output that indicates the probability of the call being genuine or fake using a sigmoid activation function. The model is optimized with the Adam optimizer, constructed using binary cross-entropy as the loss function and assessed according to correctness. A validation set was used to evaluate performance during the training phase, which comprised several epochs with a predetermined batch size. Lastly, a test set is used to assess the model's efficacy and provide an accuracy score for assessing the classification performance

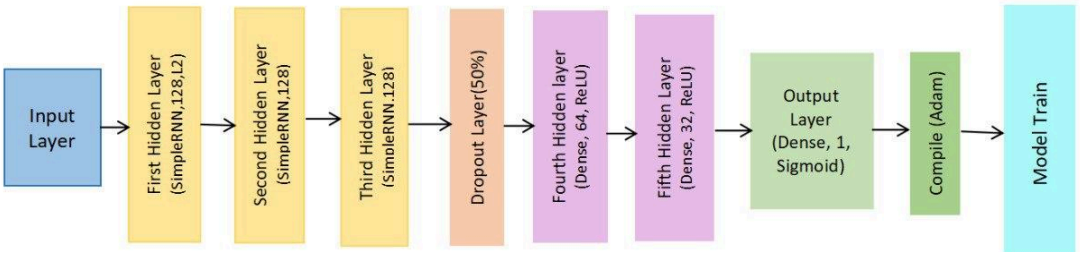


Figure 4.2.1.2: Working Flow of RNN for my work

Long Short-Term Memory (LSTM)

An LSTM-based Recurrent Neural Network (RNN) is essential for carefully

examining consecutive text input in order to categorize fraudulent phone calls. This complex design uses dropout regularisation to strengthen generalization and prevent overfitting, and it includes LSTM layers that are skilled at capturing complex sequential dependencies. Dense layers activated by ReLU methodically refine learned patterns, further augmenting its capabilities. The model ends with a crucial sigmoid output layer that allows for accurate binary categorization. It is thoroughly evaluated for accuracy, demonstrating its resilience in accurately identifying fraudulent calls. It was trained rigorously with binary cross-entropy loss and optimized using Adam.

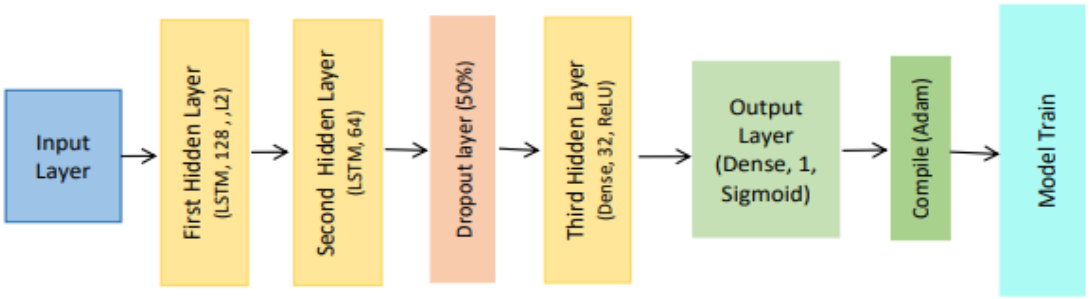


Figure 4.2.1.3: Working Flow of LSTM for my work

Bidirectional Long Short-Term Memory (BiLSTM)

An efficient method for classifying fraudulent phone calls involves using a Bidirectional Long Short-Term Memory (BiLSTM) Recurrent Neural Network (RNN) to analyze sequential text input in detail. To improve generalization and reduce overfitting, dropout regularization is incorporated into this sophisticated design. Its many BiLSTM layers enable it to capture complex sequential dependencies in both forward and backward contexts with ease. The model's capacity is increased by the dense layers that ReLU activates, which refine the learned representations. Batch normalization layers further stabilize and speed up training. Finally, the network reaches a critical sigmoid output layer for accurate binary categorization. The binary cross-entropy loss is used for rigorous training, and Adam is used to optimize the model's performance. The model's accuracy is carefully assessed, highlighting its resilience in identifying fraudulent calls.

Figure 4.2.1.4: Working Flow of BiLSTM for my work

4.2.2 Fine-tune deep learning models using ASR Transcription Dataset

In this subsection, I will explore the impact of the fine-tuning approach in transfer learning on my deep learning models—ANN, RNN, LSTM, and BiLSTM—using transcribed datasets. The objective is to enhance the models' performance in practical scenarios by leveraging additional data through a fine-tuning process. For the experiment, the dataset has been divided into training, validation, and testing sets. In the experiment, 70% of the data are in the training set, while the remaining 30% are split equally between the validation and testing sets. This process involves freezing specific layers of the pre-trained models and training the remaining layers on the new transcribed dataset. The dataset was divided into training, validation, and testing sets. For the RNN, LSTM, and BiLSTM models, the data was reshaped to fit the input requirements. Specifically, the training and validation data were reshaped to three-dimensional arrays to match the expected input shape for RNN, LSTM, and BiLSTM layers. In these models, some layers were frozen to retain the learned features from the initial training. The remaining layers were retrained using the transcribed dataset. I have used callbacks for early stopping and learning rate reduction to enhance training efficiency and prevent overfitting. Validation data was used to monitor the performance and adjust the learning process. I have checked the accuracy of the models before and after applying fine-tuning to the separated test data.

4.2.3 Creating Web Application using Deep Learning Models and Bangla ASR

After training the deep learning models I create a web application and make a pipeline between BanglaASR and Fine-tuned ANN models. This application can detect fraud and non-fraud phone calls from recordings.

4.3 Model Evaluation

4.3.1 Model evaluation for clean conversational dataset

The trained models were evaluated using a clean conversational dataset with four deep learning architectures: ANN, RNN, LSTM, and BiLSTM. The models demonstrated high training and validation accuracies. However, the training accuracy was slightly higher than the validation accuracy, suggesting a potential issue with overfitting. Nevertheless, the overall accuracy reached 90%, indicating effective training on the dataset. Table 4.3.1.1 details the training and validation accuracy of these deep learning models.

Table 4.3.1.1: Training Accuracy vs Validation Accuracy of deep learning models

Model	Training Accuracy	Validation Accuracy
ANN	0.98	0.95

RNN	0.98	0.93
LSTM	0.98	0.94
BiLSTM	0.98	0.91

Training and Validation Accuracy of Deep-learning Models:

Represents the training and validation accuracy of the deep learning models in Figure 4.3.1.1, 4.3.1.2, 4.3.1.3, 4.3.1. with the number of epochs on the x-axis and the accuracy and loss percentages on the y-axis.

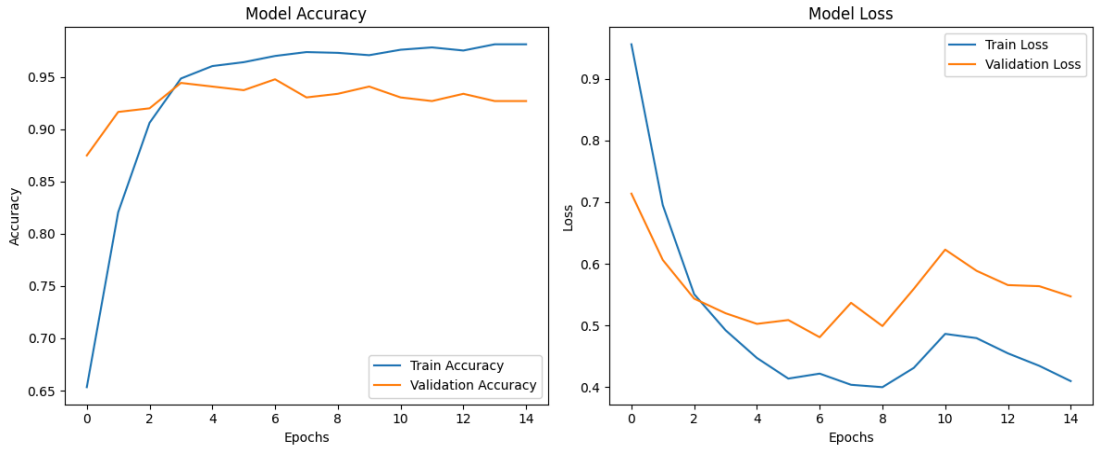


Figure 4.3.1.1: Training & validation accuracy and loss over the epochs of ANN on the clean conversational dataset

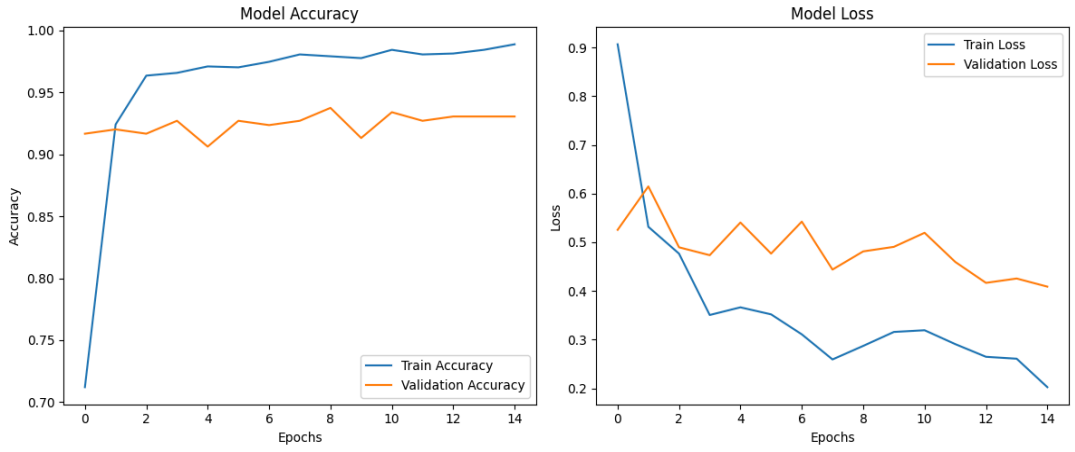


Figure 4.3.1.2: Training & validation accuracy and loss over the epochs of RNN on the clean conversational dataset

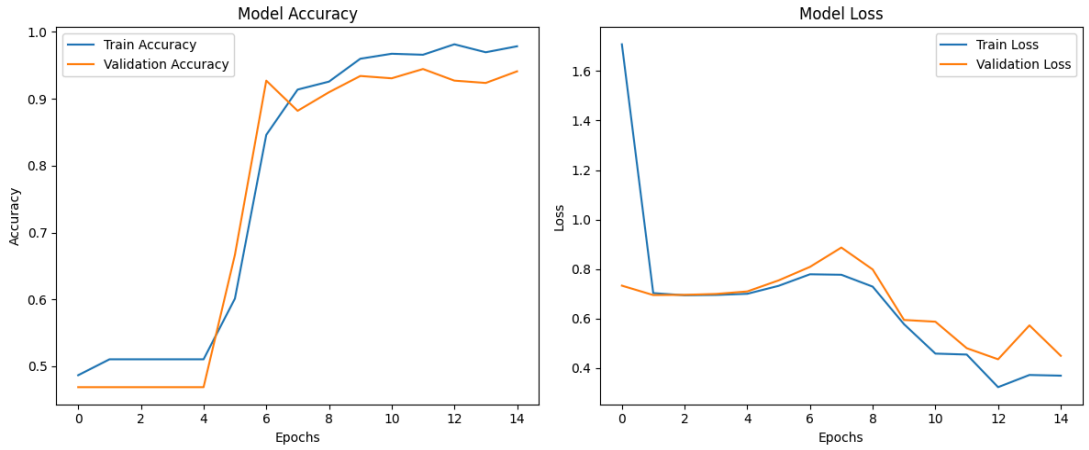


Figure 4.3.1.3: Training & validation accuracy and loss over the epochs of LSTM on the clean conversational dataset.

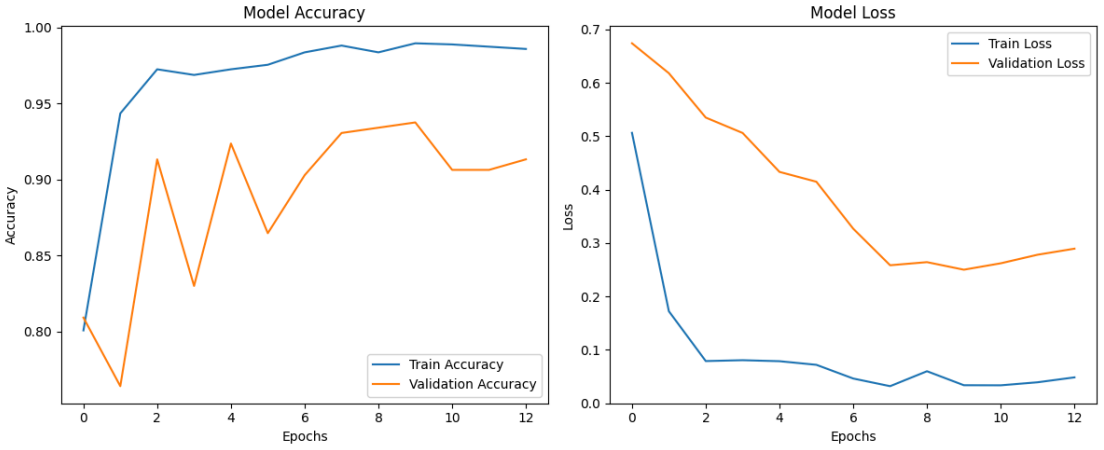


Figure 4.3.1.4: Training & validation accuracy and loss over the epochs of BiLSTM on the clean conversational dataset

4.3.2 Model evaluation for transcription dataset

I improved deep-learning models using a dataset that Automatic Speech Recognition (ASR) had transcribed. This approach aided in the model's ability to generalize their noise-induced performance. I have used early stopping and regularisation strategies to prevent overfitting because of the tiny size of the dataset.

Training and Validation Accuracy of the Fine-Tuning Process: This represents the training and validation accuracy of the deep learning modes, with the number of epochs on the x-axis and the accuracy and loss percentages on the y-axis.

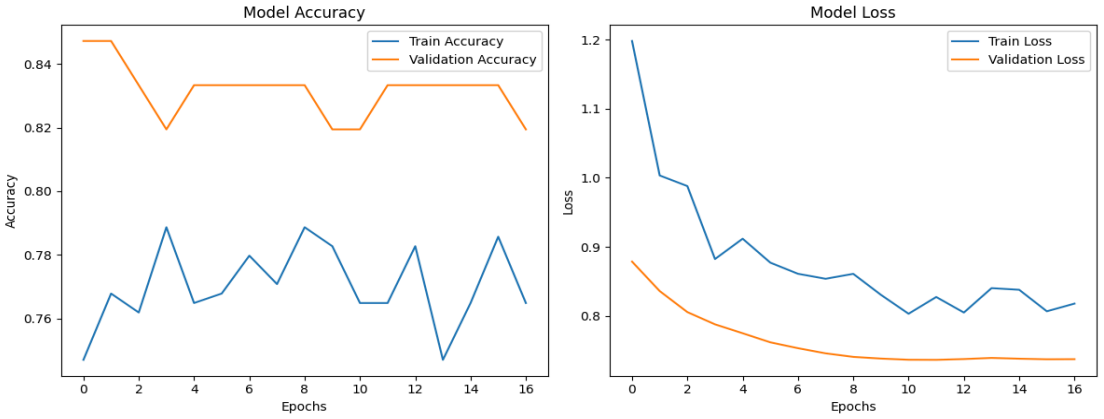


Figure 4.3.2.1: Training & validation accuracy and loss over the epochs of ANN on the transcribed dataset

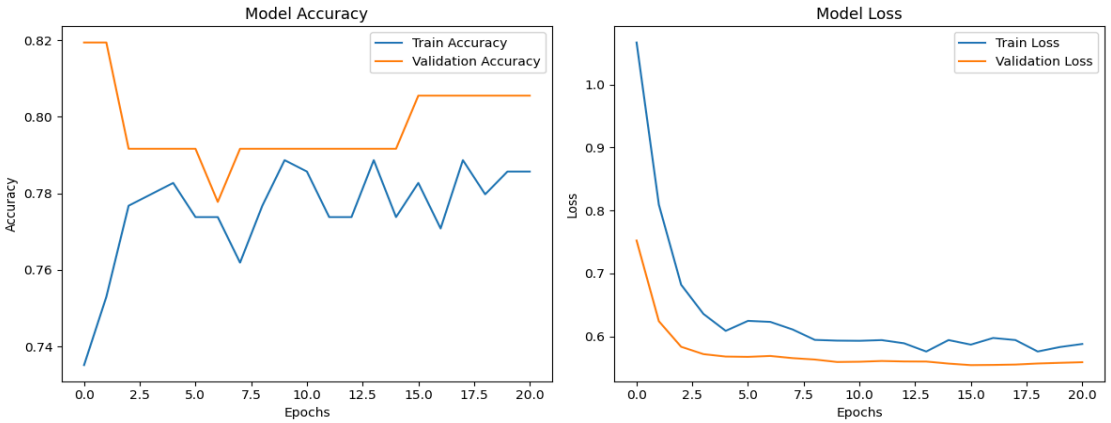


Figure 4.3.2.2: Training & validation accuracy and loss over the epochs of RNN on the transcribed dataset

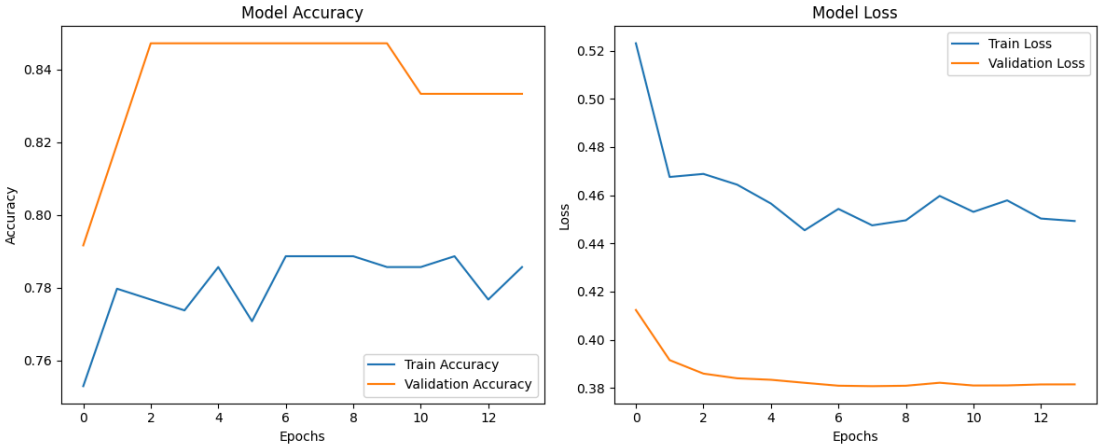


Figure 4.3.2.3: Training & validation accuracy and loss over the epochs of LSTM on the transcribed dataset

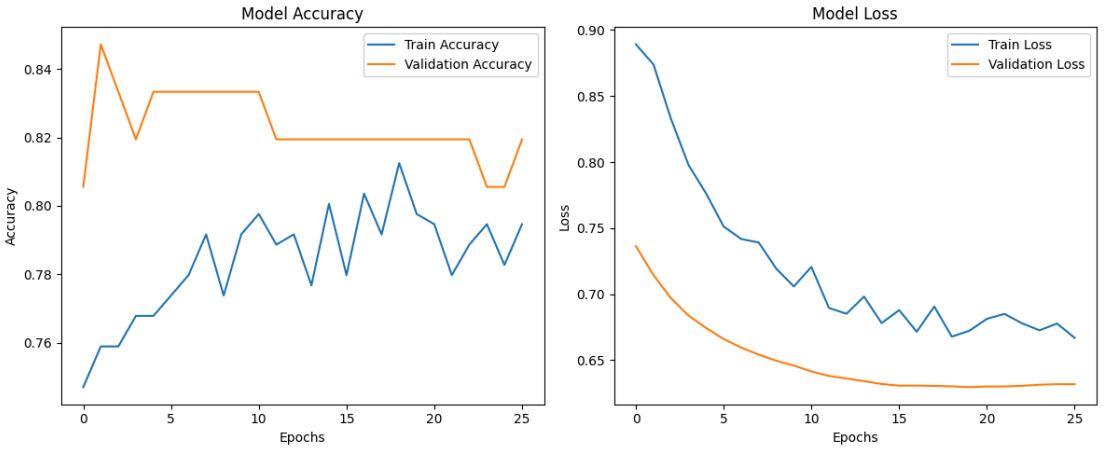


Figure 4.3.2.4: Training & validation accuracy and loss over the epochs of BiLSTM on the transcribed dataset

4.4 Summary

The implementation of my system is described in depth in this section with an emphasis on employing models on two different datasets (manual call scripts and ASR-transcribed data) and integrating them through fine-tuning. The training of deep learning models, ANN, RNN, LSTM, and BiLSTM, on these datasets, hyperparameter optimization, and performance assessment on training, validation, and test sets are covered in detail in subsection 4.2. The accuracy of the models is displayed graphically in subsection 4.3, demonstrating their usefulness and robustness in identifying fraudulent calls. Furthermore, 4.2.2's fine-tuning with ASR-transcribed data improves model performance and shows enhanced generalization skills. Although there are occasionally minor overfitting issues, overall, all models achieve high accuracy.

CHAPTER 5
RESULTS AND ANALYSIS

5.1 Overview

The performance of several models for identifying fraudulent phone calls is shown in this section, with measures like accuracy, precision, recall, and F-1 score being used to assess the models. The way the results are arranged for the two distinct datasets illustrates how well the models work with a variety of data types. I also examine how well deep learning models perform on transcribed scripts both before and after fine-tuning. These are described in detail in Section 5.3, highlighting the advantages of optimizing deep learning models for enhanced performance in identifying fraudulent phone calls.

5.2 Experimental Result

I have carried out two experiments for my investigation. One included using a clean conversation dataset to train deep learning models. Using carefully crafted scripts, this method allowed models to pick up the basic patterns and characteristics of common fraud cases. Secondly, I fine-tuned deep learning models with transcription datasets to improve performance in practical scenarios.

For evaluating models different metrics: accuracy, precision, recall, and F1-Score are used.

Accuracy: A model's accuracy can be measured by comparing its predicted and actual results to see how effectively it makes predictions. It is computed as the ratio of the total number of forecasts made to the number of accurate predictions[23].

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

Precision: Precision is implied as the measure of the correctly identified positive cases from all the predicted positive cases. The precision score indicates the overall percentage of positively predicted outcomes for all examples that were given[23].

$$\text{Precision} = \frac{TP}{(TP+FP)}$$

Recall: Recall is a measurement of correctly identified positive cases from all the actual positive cases. it gives the overall percentage of correctly positive cases from all the positive cases[23].

$$\text{Recall} = \frac{TP}{(TP+FN)}$$

F1-Score: This is the harmonic mean of Precision and Recall and gives a better measure of the incorrectly classified cases than the Accuracy Metric[23].

$$\text{F1 Score} = 2 \times \frac{Precision*Recall}{Precision+Recall}$$

5.2.1 Experiment Result of Deep Learning Models on a Conversational Dataset

The evaluation of the four deep learning models—ANN, RNN, LSTM, and BiLSTM—concentrates on their testing accuracy, precision, recall, and F1-score.

- The artificial neural network (ANN) model attained a testing accuracy of 0.95, along with a precision of 0.98, a recall of 0.94, and an F1-score of 0.95.
- The Recurrent Neural Network (RNN) model exhibited impressive performance, with a testing accuracy of 0.97. The system also achieved excellent accuracy and recall values of 0.97, resulting in an F1 score of 0.97.
- The LSTM model exhibited a little discrepancy in its precision and recall values but retained a high testing accuracy of 0.92. The accuracy achieved a value of 0.97, while the recall reached 0.87, leading to an F1-score of 0.92.
- The BiLSTM model attained a testing accuracy of 0.92, along with a precision of 0.90, a recall of 0.95, and an F1-score of 0.92.

The review emphasizes the efficacy of RNNs in collecting sequential patterns in data, rendering them especially well-suited for detecting fraudulent activities in phone call scripts. Every model demonstrates its advantages and trade-offs in categorizing fraudulent phone calls. Their accuracy and all the metrics are shown in Tables 5.2.1.1 and 5.2.1.2. The confusion matrix of these models is demonstrated in Figure 5.2.1.1.

Table 5.2.1.1: Training vs Testing Accuracy for my Conversational Dataset

Model	Training Accuracy	Testing Accuracy
ANN	0.98	0.95
RNN	0.98	0.97
LSTM	0.98	0.92
BiLSTM	0.98	0.92

Table 5.2.1.2: Precision, Recall & F1-Score for my Conversational Dataset

Model	Precision	Recall	F1-Score
ANN	0.98	0.94	0.95
RNN	0.97	0.97	0.97
LSTM	0.97	0.87	0.92
BiLSTM	0.90	0.95	0.92

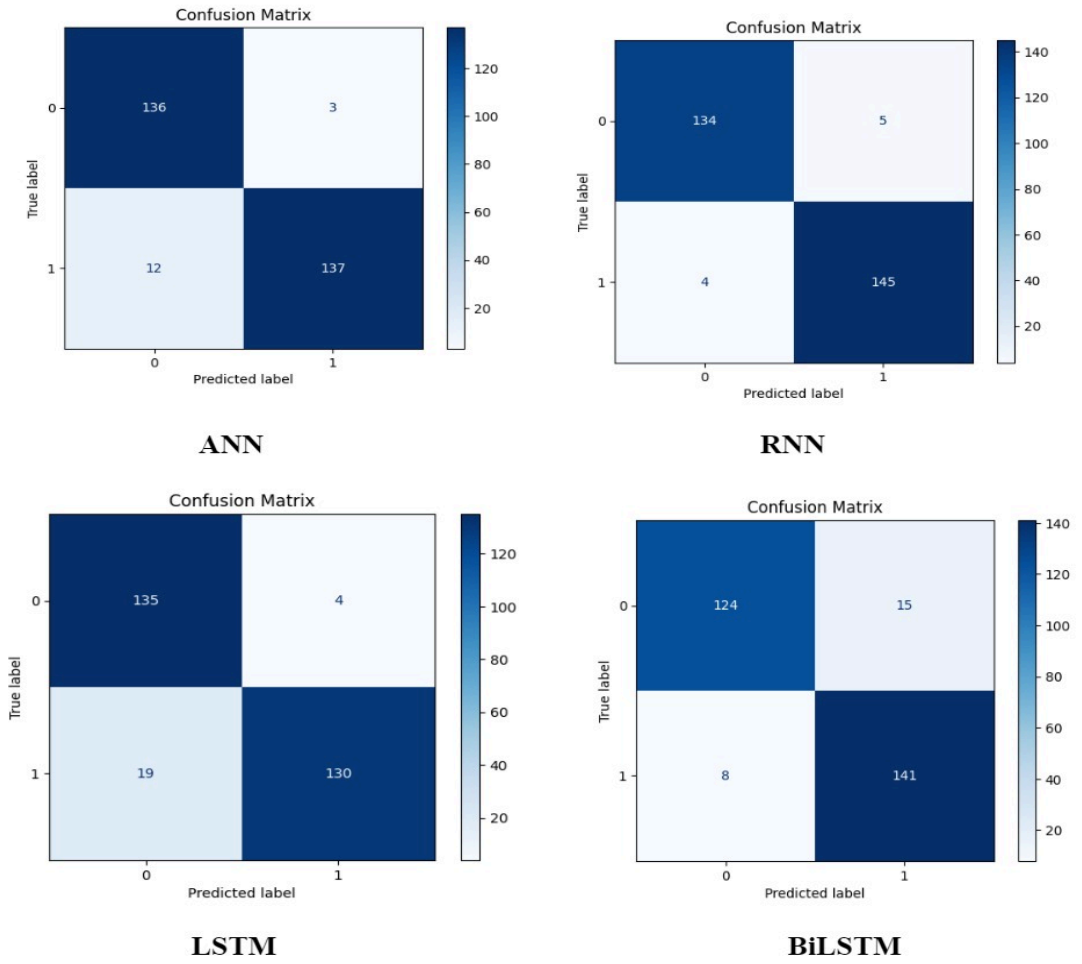


Figure 5.2.1.1: Confusion Matrix of Deep Learning Models for the clean conversational test dataset

5.2.2 Experiment Result of Fine-tuning on a Transcription Dataset

I fine-tune deep learning models on transcription datasets to improve their performance in practical scenarios. After fine-tuning it improves model performance based on accuracy and other metrics. Their performance is shown in Tables 5.2.2.1 and 5.2.2.2.

Table 5.2.2.1: Training vs Testing Accuracy for my Transcription Dataset

Model	Training Accuracy	Testing Accuracy
ANN	0.76	0.83
RNN	0.78	0.79
LSTM	0.79	0.83
BiLSTM	0.78	0.82

Table 5.2.2.2: Precision, Recall & F1-Score for my Transcription Dataset

Model	Precision	Recall	F1-Score
ANN	0.81	0.81	0.81
RNN	0.75	0.78	0.77
LSTM	0.83	0.78	0.81
BiLSTM	0.82	0.75	0.78

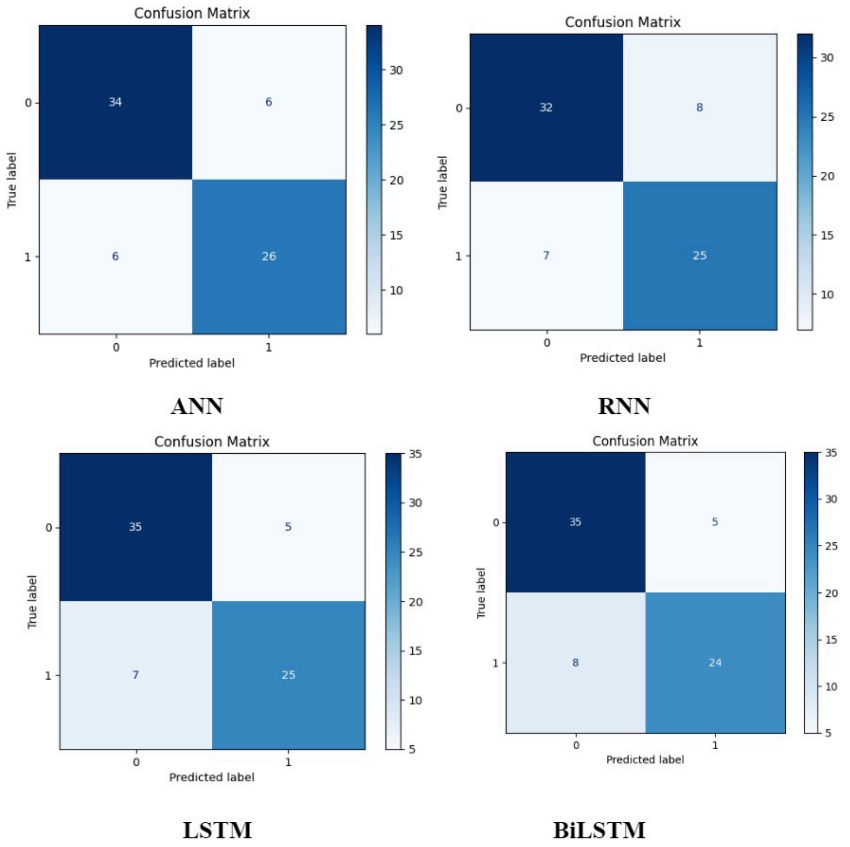


Figure 5.2.2.1: Confusion Matrix of Fine-tuned Deep Learning Models for the transcribed test dataset

5.3 Performance Analysis

The purpose of this research is to assess the performance of deep learning models in identifying fraudulent phone calls through two experiments. The WER of the transcriptions produced by the ASR system is one of the main obstacles in this categorization assignment. To overcome this problem, I have improved my pre-trained models using a dataset that the ASR had transcribed. Next, I assessed the deep learning models' performance both before and after the process of

fine-tuning.Table 5.3.1 shows the shifts in accuracy for these models.

Table 5.3.1: Accuracy Before vs After Fine-Tuning

Model	Accuracy Before Fine-tuning	Accuracy after Fine-tuning
ANN	0.77	0.83
RNN	0.65	0.79
LSTM	0.70	0.83
BiLSTM	0.75	0.82

Other important metrics like precision, recall, and F1-score are also necessary to fully evaluate the performance of the system. The comparisons are shown in Table 5.3.2.

Table 5.3.2: Performance Evaluation of Before vs After Fine-Tuning

	Before Fine-Tuning			After Fine-Tuning		
Model	Precision	Recall	F1-Score	Precision	Recall	F1-Score
ANN	0.71	0.84	0.77	0.81	0.81	0.81
RNN	0.57	0.81	0.67	0.75	0.78	0.77
LSTM	0.63	0.81	0.71	0.83	0.78	0.81
BiLSTM	0.67	0.84	0.75	0.82	0.75	0.78

After fine-tuning the model using the BanglaASR transcription dataset, these tables demonstrate a significant performance improvement. After fine-tuning four deep learning models, ANN and LSTM show a maximum accuracy of 83% with the same F1 score.

Interface of Web Application

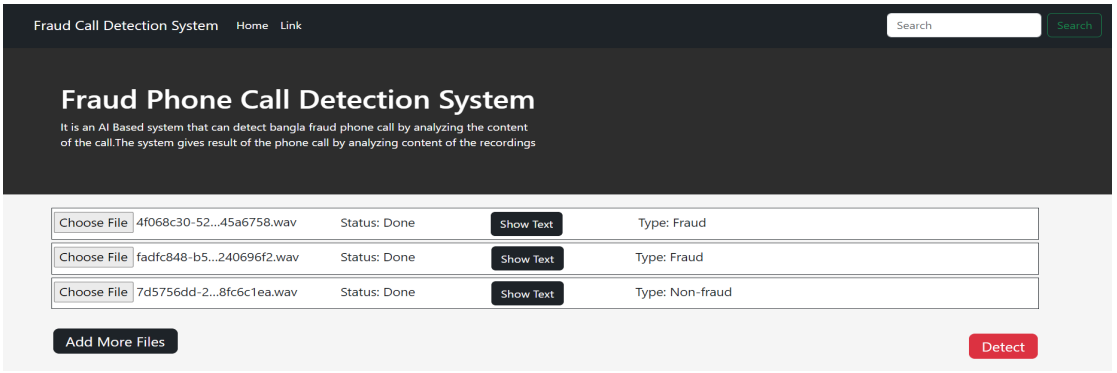


Figure 5.3.1 Web Application Interface 1

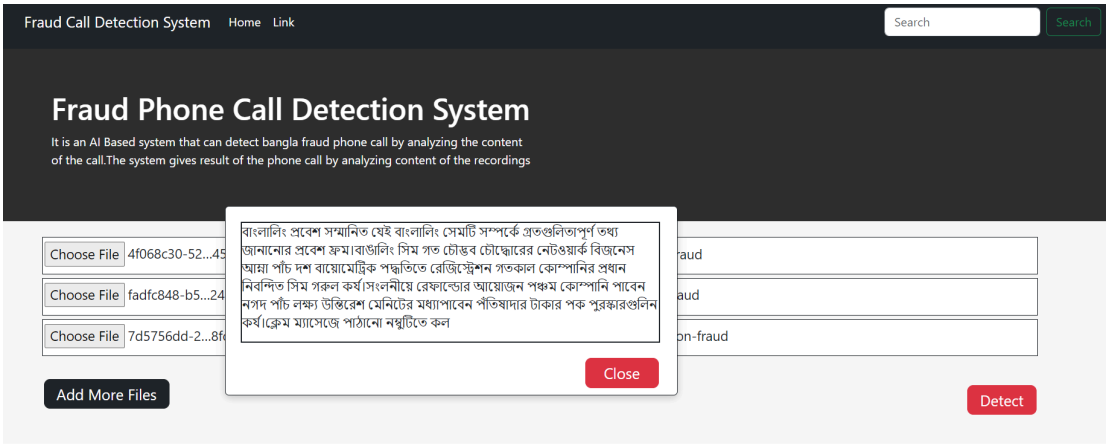


Figure 5.3.2 Web Application Interface 2

5.4 Summary

This section presents two experiments that were carried out to assess the impact of transcription errors in ASR and to study the models' performance on various datasets. Table 5.3.2, which displays the accuracy, precision, recall, and F-1 scores for every experiment, provides a summary of the findings. RNN was the top-performing model in 5.2.1.1, with a 92% accuracy rate. This outcome illustrates the efficacy of recurrent neural networks within the framework of the initial dataset. Then fine-tune these models for transcription data performance as described in section 5.2.1.2. The results show that the LSTM and ANN models are able to achieve a maximum accuracy of 83% on the transcription dataset after fine-tuning. This investigation highlights how model optimization tactics affect performance on transcription datasets.

CHAPTER 6
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

Because of technology, we can rely on the internet and telecommunication for many daily chores, which greatly simplifies our lives. This ease is not without risk, though, as con artists use these tools to trick and hurt individuals. They frequently deceive people into disclosing personal information by pretending to be someone else, and then they utilize that information to steal money and other valuables. Emotional distress and significant money losses may result from this dishonest action. My solution solves this problem by using call script analysis to identify fake phone calls. My technology can effectively discriminate between authentic and scam calls by utilizing sophisticated deep-learning algorithms. This ability is essential for assisting people in defending themselves against con artists, who are always coming up with new ways to con gullible people.

6.2 Impact on Society & Environment

This study's social impact is positive. A growing number of people are falling victim to phone call scams, which cost them a lot of money. Because of the ambiguity and fraud fear this has caused, new technologies are being used less frequently. Many people are also not tech-savvy, which makes them more vulnerable to scam calls. The use of digital money transactions has grown in the current economic environment, where many individuals are making meager salaries. Many people are easy pickings for con artists because of their dependence on technology and economic fragility. Fraudulent callers frequently take advantage of these circumstances with great success. All things considered, fraud detection technology research and application will be very advantageous to society. It helps protect people from being victims of scams, particularly those who are less tech-savvy. Future developments in this field will be welcomed by society and help create a more secure and safe online environment.

6.3 Ethical Aspects

The most concerning ethical aspects in the scope of this study are the following”

- Identity Protection of Victims
- Reputation Protection of Business Entities
- Protection of the Reputation of Individuals.

- Protection of Privacy

There is a chance that private information will be made public because the data is sensitive. I will routinely draft phone call scripts to minimize this by excluding any personal information. I will tell participants to use fictitious names and details where appropriate when they record their voices reading these scripts. Additionally, I remove company and institution names during text preparation to stop the model from picking up biases. These safeguards ensure that the ethical components of my activity are upheld and assist in preserving the four previously mentioned principles. I will try to protect sensitive data and respect ethical standards in my project by putting these safeguards in place.

6.4 Sustainability Plan

Fraud callers are always coming up with new scams and evil schemes to defraud people out of their money. As a result, after a short while, the initially trained model is probably going to perform poorly. Continuous data gathering and training are necessary to keep it effective. Another major problem is adjusting to different accents and languages. To tackle this, continuous data gathering, cooperation, and model training are needed. I may improve the robustness and efficacy of the algorithm in identifying fraudulent calls by adding a variety of language and accent data and updating it on a regular basis.

6.5 Summary

The significant effects that phone scams have on people and society are discussed in this study. While technology makes daily work easier by facilitating communication and the internet, it also makes people more vulnerable to fraudulent activities, such as identity theft and money laundering. This has caused fewer tech-savvy people to fear fraud more and have less faith in new technologies. Scams are more likely to occur in the current economic climate due to the increased reliance on digital transactions and associated issues. This research provides a protection against fraudulent calls by analyzing call scripts using advanced deep-learning techniques. It seeks to protect people from monetary and psychological harm while fostering faith in technical innovations, enhancing societal resilience and security in the digital era.

CHAPTER 7
CONCLUSION AND FUTURE WORK

7.1 Conclusions

This research proposes an innovative approach using content analysis to identify fraudulent phone calls. ANN, RNN, LSTM, and BiLSTM are the four deep-learning models that are used. A technique is provided to enhance the performance in cases of noisy transcribing. Detailed result analysis and techniques for optimizing the deep learning models to improve transcription accuracy are also explored. The accuracy of all models and fine-tuning effects were also briefly discussed. This all-inclusive system uses cutting-edge deep learning algorithms to attempt to offer a practical solution for detecting fraudulent phone calls. The work presents a particular method targeted at improving model performance to meet difficulties caused by noisy transcription circumstances. This method is essential for accurately identifying fraudulent calls even in the presence of faulty data. These approaches are accompanied by a thorough result analysis that provides insights into the efficacy of each model and highlights tactics for maximizing transcribing accuracy. The impact of fine-tuning on model performance is also discussed, highlighting the importance that fine-tuning plays in enhancing overall robustness and accuracy. By integrating these components into a cohesive system, the research provides a comprehensive solution for detecting fraudulent activities in phone calls.

7.2 Further Suggested Works

I created a web application that uses a powerful NLP model to transcribe recordings in order to identify fraudulent phone calls. Real-time fraud detection with timely user alerts to stop financial losses and privacy violations is the primary objective. In the future, I want to expand this feature to a real-time Android app by using the NLP model and giving consumers immediate fraud alerts on their smartphones. With the ability to react proactively to possible fraud efforts from their smartphones, consumers will be able to respond to attempts at fraud with more security and convenience.

7.3 Limitations

The absence of previous research and available datasets makes it difficult to detect fraudulent phone calls in Bangla language content analysis using Natural Language Processing (NLP). At now, there is a conspicuous lack of publicly available datasets that are specially designed to kickstart research in this area. The lack of ground data makes it more difficult to gather the necessary situations for training and testing fraud

detection systems in Bangla. Furthermore, it is difficult to expand on current frameworks or contrast novel ideas with tried-and-true systems because there hasn't been any prior research in this field. These drawbacks highlight the necessity of large-scale data gatherings as well as innovative research projects to establish the groundwork for efficient fraud detection systems customized for the Bangla language environment.

REFERENCES

[1] Jabbar, M. A., & Suharjito, S. B. , “Fraud detection call detail record using machine learning in telecommunications companies” . *Adv. Sci. Technol. Eng. Syst. J*, vol. *5*, pp. 63-69, July 2020.

[2] Xing, J., Yu, M., Wang, S., Zhang, Y., & Ding, Y. , “Automated fraudulent phone call recognition through deep learning.” , *Wireless Communications and Mobile Computing*, , 1-9, *2020*.

[3] Zhao, Q., Chen, K., Li, T., Yang, Y., & Wang, X., “Detecting telecommunication fraud by understanding the contents of a call.” , *Cybersecurity*, vol. *1*, pp. 1-12, July 2018

[4] Moussavou Boussougou, M. K., & Park, D. J. , “Attention-Based 1D CNN-BiLSTM Hybrid Model Enhanced with FastText Word Embedding for Korean Voice Phishing Detection. ” , *Mathematics*, Article No. *11*(14), 3217, July 2023

[5] Kumar, P. K., Ray, S., Kumarasankaralingam, L., Ramamoorthy, A., Kumar, P., & Dutta, A., “Detecting Fraud Calls vis-à-vis Natural Language Processing”, *2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)* , pp. 457-462 , May 2024.

[6] Naidu, D. , "Voice analysis system for detection of vishing using deep learning." *International Journal of Health Sciences*, May 2022.

[7] Kale, N., Kochrekar, S., Mote, R., & Dholay, S. , “Classification of fraud calls by intent analysis of call transcripts” , *12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1-6 , July 2021

[8] Natarajan, A., Kannan, A., Belagali, V., Pai, V. N., Shettar, R., & Ghuli, P. , “ Spam detection over call transcript using deep learning” . In *Proceedings of the Future Technologies Conference (FTC) 2021, Volume 2* (pp. 138-150). Springer International Publishing, 2022.

[9] Ezzat, S., El Gayar, N., & Ghanem, M. M. (2012). Sentiment analysis of call centre audio conversations using text classification. *International Journal of Computer Information Systems and Industrial Management Applications*, *4*(1), 619-627.

[10] Özlan, B., Haznedaroğlu, A., & Arslan, L. M, “Automatic fraud detection in call center conversations”, *27th Signal Processing and Communications Applications Conference (SIU)*, pp. 1-4, April 2019

[11] Mawgoud, A. A., Abu-Talleb, A., & Tawfik, B. S, “A Holistic Neural Networks Classification for Wangiri Fraud Detection in Telecommunications Regulatory Authorities. ” , *International Conference on Advanced Machine Learning Technologies and Applications*, pp. 175-183 , Cham: Springer International Publishing, March 2021.

[12] Hong, B., Connie, T., & Goh, M. K. O, " Scam Calls Detection Using Machine Learning Approaches ", *11th International Conference on Information and Communication Technology (ICoICT)* , pp. 442-447 , August 2023
[13] Elizalde, B., & Emmanouilidou, D , "Detection of Robocall and Spam Calls using Acoustic Features of Incoming Voicemails. ", AIP Publishing, *Proc. Mtgs. Acoust,* November 2021
[14] LeCun, Y., Bengio, Y., & Hinton, G. "Deep learning". *nature*, vol. *521*(7553), pp. 436-444, May 2015.
[15] Mostafa, F. A., Elrefaei, L. A., Fouda, M. M., & Hossam, A. , "A survey on AI techniques for thoracic diseases diagnosis using medical images" . *Diagnostics*, no. *12*(12), p.3034, December 2022.
[16] GeeksforGeeks, available at <<https://www.geeksforgeeks.org/introduction-to-ann-set-4-network-architectures/>>, last accessed on 28-06-2024 at 01.00 PM.
[17]GeeksforGeeks, available at << https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/>>, last accessed on 28-06-2024 at 02.00 PM.
[18] GeeksforGeeks, available at << https://www.geeksforgeeks.org/understanding-of-lstm-networks/>>, last accessed on 28-06-2024 at 02.30 PM.
[19] GeeksforGeeks, available at << https://www.geeksforgeeks.org/bidirectional-lstm-in-nlp/>>, last accessed on 28-06-2024 at 03.00 PM.
[20] Bangladesh Bank, available at <<https://www.bb.org.bd/en/index.php/financialactivity/mfsdata/>>, last accessed on 20-06-2024 at 07.10 PM
[21] IBM, available at <<https://www.ibm.com/topics/fine-tuning?fbclid=IwZXh0bgNhZW0CMTAAAR0GLfE02vqMnIgk0oUKrRhgQwz9CUmjGsydVzpYfLeaWBWCxkVfUuvM3-8_aem_7MG5xkwNIh75D6N2CmczKQ/>>, last accessed on 21-06-2024 at 8.00 PM
[22] Hugging Face, available at <<https://huggingface.co/bangla-speech-processing/BanglaASR/>>, last accessed on 21-06-2024 at 9.00 PM
[23] Medium, available at <<https://medium.com/analytics-vidhya/accuracy-vs-f1-score-6258237beca2/>> last accessed on 21-06-2024 at 10.00 PM

[24] GeekforGeeks, available at <<https://www.geeksforgeeks.org/transfer-learning-with-fine-tuning-in-nlp/>> last accessed on 21-06-2024 at 10.00 PM

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

Title: TELECOMMUNICATION FRAUD DETECTION BY ANALYZING THE CONTENT OF A BANGLA PHONE CALL
Student ID: 203-15-14475

CO Description for FYDP

CO	CO Descriptions	PO
	Phase -I	
CO1	Integrate recently gained and previously acquired knowledge to identify a telecommunication fraud phone call detection for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish this goals for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
	Phase -II	
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**

CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**
CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**
CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

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

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

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of knowledge required**	**Yes**	**CO1, CO2, CO3, CO5, CO6, CO7 and CO8**	(K1) define a systematic, theory-based understanding of the natural sciences applicable to the discipline, which information we used to build our knowledge on different techniques and proposed some preprocessing and model in our 3.2 Proposed Methodology section as well as for final defense project design. (K2) define conceptually based on mathematics, numerical analysis, statistics, and formal aspects of computer and information science to support analysis and modeling applicable to the discipline which we used for our table 2.3.1 for comparison between existing works and in table 3.4.2.1 for estimating the cost for our project in 3.4, 3.5 Project Management and Financial Analysis. In chapter 4 Table 4.3.1.1 and 4.3.2.1 describe the training and Validation accuracy of find in deep learning models. This project demonstrates fundamental engineering (K3) principles by employing deep neural networks, data augmentation techniques, and different deep learning model architectures for natural language processing and classification tasks. The project demonstrates specialist knowledge (K4) by conducting The project demonstrates specialist knowledge (K4) by conducting fine-tuning of pre-trained deep learning model on a transcribed dataset by Bangla ASR.	**Page no:** [13-21] **Section:** [3.2] **Page no:** [11-12,22-23, 28-32] **Section:** [2.3,3.4,4.3] **Page no:** [13-21] **Section:** [3.2] **Page no:** [25-28] **Section:** [4.2]

				The project applies engineering practice & design (K5) by the figure of process of experiments. The project addresses engineering practice & technology (K6) by employing deep learning models.	**Page no:** [12] **Section:** [2.4] **Page no:** [25-28] **Section:** [4.2]
				This project ensures to K8 (Research Literature) by synthesizing insights from recent studies, to detect fraudulent phone call using deep learning and NLP techniques, showcasing a comprehensive understanding of current methodologies.	**Page no:** [7-12] **Section:** [2.2, 2.3]
2.	**EP2: Range of Conflicting Requirements**	**Yes**	**CO2, and CO7**	This project addresses EP-2 by recognizing the hurdles in telecommunication fraud detection, including limitations of traditional methods and the complexities in fraud phone call detection. Through comparative analysis, it confronts challenges in understanding spatial distributions, offering insights of different approaches.	**Page no:** [12] **Section:** [2.4]
3.	**EP3: Depth of analysis required**	**Yes**	**CO2, and CO6**	This project addresses EP-3 by meticulously comparing experimental outcomes, highlighting fine-tuning of deep learning models to improve accuracy in fraud phone call detection.	**Page no:** [33-36] **Section:** [5.2]
4.	**EP4: Familiarity of Issues**	**Yes**	**CO8**	This project's interdisciplinary approach extends beyond computer science and engineering, impacting telecommunication sector by external use of new system to detect fraud phone call contributing to advancements of telecomunication sector.	**Page no:** [3-4] **Section:** [1.5]

5.	**EP5: Extends of application codes**	**No**	**CO5**	N/A	**N/A**
6.	**EP6: Extends of stakeholders involved and conflicting requirements**	**No**	**CO8**	N/A	**N/A**
7.	**EP7: Interdependence**	**Yes**	**CO5**	This project's comprehensive approach addresses high-level problems by integrating various components across data collection, data augmentation, data preprocessing and proposed methodology, ensuring a holistic solution to complex challenges in telecommunication sectors which ensures EP-7.	**Page no:** [13-21] **Section:** [3.2]

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

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	**Yes**	**CO11**	Our project utilizes diverse resources such as high-performance computing infrastructure, GPUs, deep learning frameworks, annotated datasets, and ethical considerations to ensure systematic research and contribute to advancements in telecommunication fraud detection.	**Page no:** [39-40] **Section:** [6.3]
2.	**EA2: Level of interaction**	**No**		N/A	**N/A**

3.	**EA3: Innovation**	**No**		N/A	**N/A**
4.	**EA4: Consequences for society and the environment**	**Yes**		This project contributes to society by detecting fraud phone call in telecommunication sector through advanced fraud detection method, while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines.	**Page no:** [39-40] **Section:** [6.1,6.2,6.4]
5.	**EA-5: Familiarity**	**Yes**		This project expands upon existing research by examining a novel approach in fraud phone call detection demonstrated through preliminary terminologies and a comprehensive comparative analysis, offering new insights into the field.	**Page no:** [7,11-12] **Section:** [2.1, 2.3]

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

SN	COs	Attainment	Justification	References
1	**CO4**	**Yes**	This project addresses CO4 by integrating effective project management and financial oversight, ensuring meticulous planning, resource allocation, and budget estimation for optimal resource utilization throughout the research lifecycle.	**Page no:** [22-23] **Section:** [3.4]
2	**CO9**	**Yes**	The project demonstrates adherence to ethical principles by prioritizing individual privacy, obtaining informed consent, and transparently documenting the research process, ensuring responsible knowledge dissemination and societal well-being through the ethical application of this fraud detection system.	**Page no:** [39-40] **Section:** [6.3]
3	**CO10**	**No**	**N/A**	**N/A**

4	**CO12**	**Yes**	The project's dedication to continuous learning (CO12) and adaptation within the dynamic technological landscape is reflected in its comprehensive data collection, rigorous statistical analysis, meticulous methodology development, and thorough experimental results and analysis, showcasing a commitment to staying updated and refining techniques to address modern challenges.	**Page no:** [13-24, 25-28] **Section:** [3.2, 3.3, 3.4, 3.5, 4.2]

TELECOMMUNICATION FRAUD DETECTION BY ANALYZING THE CONTENT OF A BANGLA PHONE CALL

ORIGINALITY REPORT

10%	8%	6%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	2%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
3	medium.com Internet Source	<1%
4	www.ibm.com Internet Source	<1%
5	www.techscience.com Internet Source	<1%
6	repositorio.usm.cl Internet Source	<1%
7	gala.gre.ac.uk Internet Source	<1%
8	mdpi-res.com Internet Source	<1%
9	link.springer.com Internet Source	<1%