

# Comparative Analysis of Different Deep Learning Models to Classify Infectious Diseases from Chest Radiological Images

*by Md Hasibur Rahman*

---

**Submission date:** 11-Jan-2025 05:40PM (UTC+0600)

**Submission ID:** 2552353055

**File name:** Hasibur Rahman.pdf (9.10M)

**Word count:** 13242

**Character count:** 62937

## **Emotion Detection from Bangla Text Using Seven Emotion Classes**

By  
Md Toufiqur Rahman Plabon  
Student ID: 211-15-14579

### **FINAL YEAR DESIGN PROJECT REPORT**

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

#### **Supervised By**

**Mr. Md. Sazzadur Ahamed**

**Assistant Professor**

Department of Computer Science & Engineering

Daffodil International University

#### **Co-Supervised By**

**Ms. Umme Ayman**

**Lecturer**

Department of Computer Science & Engineering

Daffodil International University



## **DAFFODIL INTERNATIONAL UNIVERSITY**

**Dhaka,Bangladesh**

**November 15, 2024**

## **APPROVAL**

This Project titled “**Emotion detection from Bangla text**”, submitted by **Md. Toufiqur Rahman Plabon** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **15-12-2024**.

## **BOARD OF EXAMINERS**

---

**Name**  
**Board Chairman**  
Designation, Department of CSE,  
FSIT Daffodil International  
University

---

**Name**  
**Internal Examiner 1**  
Designation, Department of CSE,  
FSIT Daffodil International  
University

---

**Name**  
**Internal Examiner 2**  
Designation, Department of CSE,  
FSIT Daffodil International  
University

---

**Name**  
**External Examiner**  
Designation, Department of CSE,  
FSIT Daffodil International  
University

## **DECLARATION**

We hereby declare that this project has been done by us under the supervision of **Mr. Md. Sazzadur Ahamed, Assistant Professor, Department of Computer Science and Engineering, Daffodil International University.** We 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:**

---

#### **Mr. Md. Sazzadur Ahamed**

Assistant Professor  
Department of Computer Science and Engineering  
Daffodil International University

### **Co-Supervised by:**

---

#### **Ms. Umme Ayman**

Lecturer  
Department of Computer Science and Engineering  
Daffodil International University

### **Submitted by:**

3

---

#### **Md. Toufiqur Rahman Plabon**

Student ID: 211-15-14579  
Department of Computer Science and Engineering  
Daffodil International University

## ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

1

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the Final Year Design Project (FYDP) successfully.

We are grateful and wish our profound indebtedness to **Mr. Md. Sazzadur Ahamed**, Assistant Professor, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of Deep learning to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

3

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

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

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

## ABSTRACT

The focus of this thesis is the categorization of seven emotional states: anger, happiness, surprise, fear, sadness, confusion, and disgust. Additionally, a emotion detection study was carried out in Bangla text. This capability of automatically detecting emotions in text has grown in value with the expansion of digital communication. It can be used in social media analysis, consumer feedback, and mental health assistance. However, the complexity in Bangla morphologically and syntactically rich language are often missed by the existing emotion recognition methods that are mostly built over high resource languages like English. The aim of the study is to develop a classification model that will identify emotion from a single Bangla text sample with high accuracy in overcoming these limitations. A dataset was designed and annotated for the study, and then pre-processing techniques like tokenization, normalization, stemming applied specifically for Bangla. Efficiency in handling classification was tested by several machine learning and deep learning models. Model performance for each category of emotion has been presented with the help of key evaluation measures: precision, recall, F1-score. Confusion matrix is also shown in this paper.

**Keywords:** Emotion, Bangla Text, Social Media, Machine Learning, Deep Learning, Accuracy

# Table of Contents

<b>Approval</b>	i
<b>Declaration</b>	ii
<b>Acknowledgements</b>	iii
<b>Abstract</b>	iv
<b>3 List of Figures</b>	vii
<b>List of Tables</b>	viii
<b>1 Introduction</b>	1-5
1.1 Introduction.....	1,2
1.2 Motivation .....	2,3
1.3 Objectives .....	3
1.4 Methodology .....	3
1.5 Project Outcome .....	4
1.6 Organization of the Report .....	4
<b>2 Background</b>	5-14
2.1 Introduction.....	5
2.2 Literature Review .....	5-13
2.3 Gap Analysis .....	13-14
2.4 Summary .....	14
<b>3 Research Methodology</b>	15-31
3.1 Methodology/Requirement Analysis & Design Specification.....	15
3.1.1 Overview .....	15
3.1.2 Proposed Methodology/ System Design.....	16

**Table of Contents**

<b>Table of Contents</b>		
3.2	Detailed Methodology and Design .....	16-29
3.2.1	Data Collection.....	16
3.2.2	Data Cleaning .....	17
3.2.3	Data Preprocessing.....	19-22
3.2.4	Logistic Regression.....	22
3.2.5	Random Forest Classifier.....	23
3.2.6	Multinomial Naïve Bayes.....	24
3.2.7	Support Vector Machine.....	24
3.2.8	Gradient Boosting Model.....	25
3.2.9	CNN.....	25-26
3.2.10	Bangla Bert.....	27
3.2.11	RNN.....	27-28
3.2.12	DNN.....	28-29
3.3	Project Plan.....	29-30
3.4	Task Allocation.....	31
3.5	Summary.....	31
<b>4</b>	<b>Implementation and Results</b>	<b>32-45</b>
4.1	Environment Setup.....	32
4.2	Performance Analysis .....	33-34
4.3	Results and Discussion .....	35-44
4.4	Summary .....	45
<b>5</b>	<b>Engineering Standards and Design Challenges</b>	<b>47-53</b>
5.1	Compliance with the Standards .....	47
5.1.1	Communication Standards .....	47
5.2	Impact on Society, Environment and Sustainability .....	47-48
5.2.1	Impact on Life.....	48
5.2.2	Impact on Society & Environment.....	48
5.2.3	Ethical Aspects.....	48
5.2.4	Sustainability Plan.....	48
5.3	Project Management and Financial Analysis.....	49
5.4	Complex Engineering Problem .....	50
5.4.1	Complex Problem Solving .....	50-52
5.4.2	Engineering Activities.....	53
5.5	Summary .....	53
<b>6</b>	<b>Conclusion</b>	<b>54-55</b>
6.1	Summary .....	54
6.2	Limitation .....	54

6.3 Future Work .....	49 54-55
<b>References</b>	<b>56-57</b>

## LIST OF FIGURES

<b>FIGURES</b>	<b>PAGE NO</b>
Figure 1.1.1 The 2D valence arousal model	2
Figure 3.1.2.1 Methodology Process	16
Figure 3.2.4.1 Logistic Regression Architecture	23
Figure 3.2.5.1 Random Forest Model Architecture <span style="background-color: #c0c0c0; padding: 2px 5px;">7</span>	24
Figure 3.2.9.1 CNN Architecture <span style="background-color: #c0c0c0; padding: 2px 5px;">7</span>	26
Figure 3.2.10.1 Bangla Bert Architecture	27
Figure 3.2.11.1 RNN Architecture	28
Figure 3.2.12.1 DNN Architecture <span style="background-color: #c0c0c0; padding: 2px 5px;">13</span>	29
Figure 4.3.1.1 Logistic Regression Confusion Matrix	37
Figure 4.3.2.1 Random Forest Model Confusion Matrix <span style="background-color: #c0c0c0; padding: 2px 5px;">1</span>	38
Figure 4.3.3.1 Multinomial Naïve Bayes Confusion Matrix	39
Figure 4.3.4.1 SVM confusion matrix <span style="background-color: #c0c0c0; padding: 2px 5px;">1</span>	40
Figure 4.3.5.1 Gradient Boosting Classifier Confusion Matrix <span style="background-color: #c0c0c0; padding: 2px 5px;">1</span>	41
Figure 4.3.6.1 CNN model Confusion Matrix <span style="background-color: #c0c0c0; padding: 2px 5px;">70</span>	42
Figure 4.3.6.2 CNN model Training and Validation Accuracy	42
Figure 4.3.6.1 CNN model Training and Validation Loss <span style="background-color: #c0c0c0; padding: 2px 5px;">1</span>	42
Figure 4.3.7.1 Bangla Bert Confusion Matrix	43
Figure 4.3.8.1 RNN Model Confusion Matrix <span style="background-color: #c0c0c0; padding: 2px 5px;">33</span>	44
Figure 4.3.8.2 RNN Model Training and Validation Accuracy	44
Figure 4.3.8.3 RNN Model Training and Validation Loss	44

Figure 4.3.9.1 DNN Model Confusion Matrix	45
8 Figure 4.3.9.2 DNN Model Training and Validation Accuracy	45
8 Figure 4.3.9.3 DNN Model Training and Validation Loss	45

## **LIST OF TABLES**

<b>TABLES</b>	<b>PAGE NO</b>
Table 2.2.1 Summary of literature review	10-13
Table 3.2.1.1 Sample Collected Dataset	16-17
Table 3.2.2.1 The Final Dataset	18
Table 3.2.3.1 Cleaned Dataset Table	19
Table 3.2.3.2 Tokenized Dataset Table	20
Table 3.3.1 Project Plan	29-30
Table 4.3.1 Performance Analysis Metrics	35-36
Table 5.3.1 Primary Budget	49
Table 5.3.2 Alternate Budget	49
Table 5.4.1.1 Mapping with complex problem solving.	50
Table 5.4.1.2 Mapping with knowledge Profile (EP1)	51
Table 5.4.1.3 Mapping with knowledge Profile (EP3)	51
Table 5.4.1.4 Mapping with knowledge Profile (EP7)	52
Table 5.4.2.1 Mapping with complex engineering activities	53

## CHAPTER 1

### INTRODUCTION

#### 1.1 Introduction

Psychologists are studying in “Human Emotion” from the very beginning. It has been an intriguing topic to explore and analyze. Emotion refers to “Subjective Experience, Psychological Response, Behavioral Response.” “Behavioral Response” indicates how the emotions are expressed. “Physiological arousal, expressive behaviors, and conscious experience” are involved with human emotion said by David G [1]. Human interacts with emotions each and every moment. They express emotions with a smile or facial expressions, verbal speech, hand sign or sign language and written text. By the advancement of “Technology” we have to come to this point that we interact with global world through a common platform that is called social media. People are expressing themselves simultaneously and spontaneously. They can express their feelings through video, audio or written text. Emotion detection is easy from video or audio more than written text. Since a written text can convey a wide range of emotions, including anger, sadness, fear, confusion, happiness, disgust, surprise, positivity or negativity. Texting one to one is confidential and it doesn't have a broad impact. On the other hand, commenting on a post or reviewing on a product can have an influence over others. Millions of data are being generated every day and are guiding us in many ways. These data are being generated based on our emotions. Detecting the right emotion from the data is a significant subject in this era. There has been extensive research conducted on detecting emotions from English corpus. But detecting emotion from Bangla is very few and the accuracy is not satisfactory. But Bengali is the Seventh most spoken native language and seventh most spoken language by the total number of speakers of the world. Millions of people are expressing their emotion though Bengali frequently in their daily life. It is a matter of concern whether their emotion is conveyed to others correctly or not.

This paper aim is to detect ‘happy, sad, angry, confused, fear, surprise, disgust. Data has been collected from comments on social media, reviews from e-commerce sites. Data has been annotated carefully. Then data has been preprocessed. After that we trained the established model and then tested it. The 1.1 figure describes the activation and deactivation parts of our emotions.

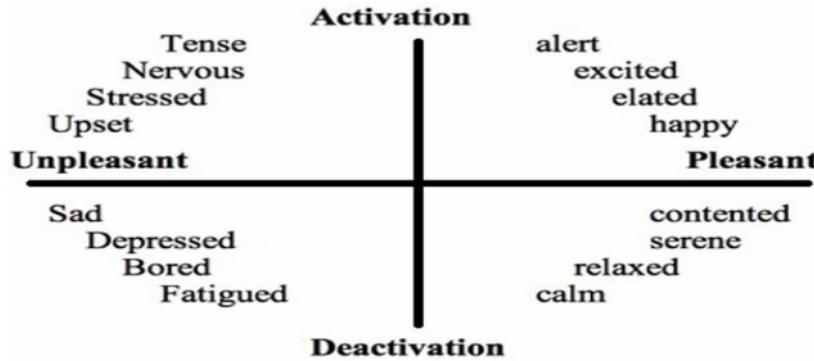


Figure 1.1 The 2D valence arousal model [2]

76

One of the most important tasks in natural language processing (NLP) is identifying emotions in textual data. This allows for applications in sentiment analysis, customer feedback analysis, mental health monitoring, and more. Emotion recognition for high-resource languages like English has advanced significantly, but because of its particular linguistic and resource limitations, emotion detection for Bangla text confronts a number of difficulties. This study is covering the issues that are mentioned below:

- Is dataset collection and annotation done correctly?
- How to preprocess the dataset?
- Is machine learning doing better or deep learning doing better?
- Which model is detecting emotion more accurately?

### 1.1 Motivation

This rapid development of social media, online communication platforms, and the increased popularity of digital content development has dramatically changed the way in which people communicate their emotions by text. Human communication is saturated with emotions that form relationships, influence decisions, and encounters. Applications such as sentiment-based recommendation systems, mental health monitoring, and consumer feedback analysis now rely heavily on this understanding of emotions. However, while emotion recognition has significantly improved for global languages like English, low-resource languages like Bangla lag far behind due to a lack of studies and datasets.

64

Bangla is the seventh most spoken language in the world and is extremely important on a social, cultural, and economic level. One of the major impediments toward exploiting the full potential of Bangla in various technical breakthroughs is the lack of reliable methods for emotion recognition in writing. Most Bangla sentiment and emotion analysis studies concentrate on

binary or coarse-grained sentiment categorization and overlook the intricacy and richness of multilabel emotions.

Since real-world text frequently conveys several emotions all at once, such as a mix of sadness and confusion or pleasure and surprise, multi-label emotion recognition is most important. This needfully complex understanding is called for in the development of systems that can better perceive human emotions. The motivation for this thesis was therefore to solve the challenges in the detection of multilabel emotion in Bangla text, a pressing need in this field. The research work will try to enhance NLP for Bangla and bridge the gap between technology and linguistic variation through proposing a comprehensive framework for the recognition and evaluation of emotions along several dimensions.

## **7** **1.2 Objectives**

The aim of this thesis is to develop a reliable and efficient model for emotion detection in Bangla text. The objective of the work is to identify from textual inputs reliably a number of co-occurring emotional states, which includes anger, happiness, surprise, fear, sadness, confusion, and disgust. This covers to the following:

Creating or selecting a collection of Bangla texts that have been annotated with these feelings.

35  
Investigation and development of the latest natural language processing and machine learning techniques for appropriate classification.

63  
Performance of the proposed model measuring using common measures such as accuracy, precision, recall, and F1-score to ensure dependability in identifying overlapping emotional categories.

## **1.3 Methodology**

In the methodology section we have discussed how we collected the data and from where we collect the data. After that how we preprocessed the data. How we extract the feature from the data. Then, we applied the machine learning and deep learning techniques on the preprocessed data. We applied different types of models so that we can see the differences and the performance.

#### **1.4 Project Outcome**

The creation of an annotated dataset for the Bangla text, including several emotional classes such as fear, sadness, confusion, anger, happiness, surprise, and disgust, can be done. pertinent research papers, techniques, and technologies employed in the industry. The annotated dataset is preprocessed through multiple steps and cross validation is ensured. So, the words that have been preprocessed is in Bangla dictionary. Implying models that can detect emotion more accurately. Every model is applied cautiously. Thus, we can come up with the best model for detection emotion for this annotated dataset. The creation of detailed assessment criteria and standards for the identification of multilabel emotions in Bangla text. This assessment could help in future the researchers.

1

#### **1.5 Organization of the Report**

Our report is divided into six sections. Chapter 1 covered the following issues: study topics, research methodology, research potential, motivation, goal, expected results, and problem description. In chapter two, we discussed nomenclature, gaps in prior studies, and summaries. Chapter 3 included the overview, study methodology, dataset, gamma correction, split, deep learning, data preprocessing, and summary. Chapter Four was supported by Model Analysis, Visualization, Performance Analysis, Evaluation Technique, and Result Discussion. Chapter five presented discussions on the sustainability strategy, the effects on society and the environment, project management and financial analysis, and complicated technical problems. Chapter 6 was closed with discussions of Conclusion, Limitations, Future work.

## CHAPTER 2

### BACKGROUND

#### 2.1 Introduction

This part helps readers understand the current level of knowledge in the topic and approaches already in use for similar projects. By analyzing related literature, authors can highlight the study's novelty or uniqueness and place its findings within the broader research environment.<sup>19</sup> Natural language processing is a part of Artificial Intelligence which deals with the interaction between computers and human language. We have used raw data. We have classified the data into seven classes. The classes are happy, sad, angry, confused, fear, surprise, disgust. There is no study have been found where these seven classes have been involved for detecting emotion. Deep learning, Machine learning approached have been taken to in this study to detect the emotion from the labeled data. Systems capable of self-improvement from experience, without explicit programming by a subset of artificial intelligence is referred as machine learning. A branch of machine learning that extracts intricate patterns from data by using multi-layered artificial neural networks is referred as deep learning. The background material sets the appropriate context for the data presented in the article. The background of the study arouses the viewer's interest in the issue raised by the findings and shows the importance of the issue.

#### 2.2 Literature Review

Examining pertinent literature enables us to identify any gaps or topics that require more research. The limitations or shortcomings of previous studies may serve as a guidance when formulating research questions or objectives that aid in filling in these gaps. This helps us make sure that their research was worth it and also furthers the field. This subject was made known through numerous languages and forms. Here are a few relevant works that can help us in empowering the concept.

S. Azmin and K. Dhar [3] paper focuses on the use of a Multinomial Naïve Bayes classifier for classifying emotions such as happy, sad, and furious from Bangla text. The relatively unexplored field of emotion recognition in Bangla text is the focus of this work. It presents a machine learning-based method for classifying Bangla comments into three emotion categories by utilizing the Multinomial Naïve Bayes classifier in conjunction with a number

of linguistic preprocessing

approaches. This dataset contains 4,200 Facebook post comments. It has been split into 3,780 training samples and 420 test samples. Emotions like happy, sad, and angry were taken into consideration.<sup>93</sup> Part-of-Speech (POS) tagging using a tagger based on Hidden Markov Model. Word n-grams are uni-, bi-, and tri-grams. The best model performed using POS tagging, preprocessing, and bigram-based TF-IDF with an accuracy of 78.6%. The model performed poorly in the "sad" class but was good in the "happy" class due to an imbalanced data distribution.

Rowshan Rahman Rushan et al. [4] paper presents a systematic approach toward the development of a Bangla text emotion detector and associates those emotions with appropriate emojis. The study presents specific objectives, which are collecting a diverse dataset, data preprocessing, and applying multiple machine learning models. It also points out the novelty of the approach, including emoji incorporation and cultural sensitivity in Bangla language emotion recognition. The databases, containing encoded emojis, sentiment labels, and text descriptions, are documented in detail. The approaches to data preprocessing such as EDA, lemmatization, and tokenization are sufficiently explained. Only three emotions are studied in the paper: happiness, sadness, and anger. It would be more complete by including a wider range-for example, fear, surprise, and disgust. Since the majority of the datasets are originating from social media platforms, they may not reflect many situations where Bangla texts are being used. Using Bangla Bert they gain 85% accuracy.

Abdullah Al Jamil , Rifat Rahman[5] analyzes the task of emotion classification using various feature extraction methods and machine learning classifiers on Bangla and other multilingual text data. For emotion-specific multilingual text, the study used Bangla. The machine learning classifiers for identifying emotions like anger, fear, joy, and melancholy in multilingual text, including Bangla, were used. 7,103 samples from the WASSA-2017 dataset which was tagged for four different classes of emotions were used. They extended this to low-resource languages with the addition of a Bangla dataset of 600 manually collected samples. Feature extraction was done on textual data using the Count Vectorization and TF-IDF techniques. The approach they proposed had an accuracy of 74.3% on the Bangla dataset.

Amit Kumar Das et.el.[6] examines the use of deep learning algorithms in order to detect

98

hate speech in Bangla social media texts. The authors have grouped the Bangla social media comments into seven categories, such as hate speech, aggressive comments, religious hatred, ethnic attack, religious comments, political comments, and suicidal comments. It would be performed using an encoder-decoder-based neural network model with an attention mechanism. The authors manually annotated 7,425 Facebook comments into seven categories, including special classes like "suicidal comments. The data was preprocessed using methods like stopword removal, tokenization, stemming, and a custom-designed Bangla Emot Module for emojis and emoticons. This paper utilized 1D Convolutional Neural Networks (CNNs) as encoders to record the text's temporal and spatial relationships investigated three decoder architectures: an attention-based RNN, GRU, and LSTM. The most accurate RNN, attention-based, was 77%.

Shahidul Islam Khan et. el.[7] proposes a traditional statistical approach to ranking terms in a document based on rarity and frequency. A methodology to record the frequency of words present within a document. The models are used in this paper are Random Forest, Linear Support Vector Machines, Naïve Bayes, and Logistic Regression. The data has been labeled into four classes and they are sadness, joy, fear, and anger. The authors have collected 600 samples manually to detect emotions in the Bangla language. The accuracy of the proposed model is 74.3%.

31

Moshiur Rahman Faisal et al. [8] addresses the challenges of emotion detection in Bengali and Banglish with regard to the complexity of multilingualism and limited language resources. This dataset is an important resource for low-resource NLP, as it contains 80,098 items for each of the six emotions: disgust, fear, joy, sadness, surprise, and anger, in both Bengali and Banglish. Comparison with transformer, DL and classical ML models set the baseline for further study. BanglaBERT performed best with a weighted F1 score of 71.30% for Bengali and 64.59% for Banglish. The linguistic and cultural variety cannot be represented by data acquired from only YouTube and Twitter. The focus on BanglaBERT and related designs may limit the exploration of other promising approaches.

Md Ataur Rahman, Md Hanif Seddiqi [9] investigates fine-grained emotion classification in Bangla, a low-resource language, focusing on six emotions: sadness, happiness, disgust, surprise, fear, and anger. The corpus contains 6,314 manually annotated Facebook group remarks with a sociopolitical focus. Five methods are explored in depth: K-Means Clustering, Naïve Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), and k-

**Nearest Neighbor (k-NN).** The best performance was given from the nonlinear SVM (RBF kernel) with a performance of 52.98% accuracy, F1-macro: 0.3324. performance benchmarks were set. Therefore, different preprocessing methods including TF-IDF representation, feature vectors, unigram and n-gram, were tried and tested. Feature extraction included POS tagging, considering tags important for emotions, such as nouns and adjectives. The best model has shown potential for improvement by having a low F1-macro of 0.3324 and modest accuracy of 52.98%, considering the modifications made.

Tahmina Akter et. el. [10] this paper includes some of the ML and DL models for emotion recognition. They are **k-NN**, **Decision Tree (DT)**, **Random Forest (RF)**, **SVM**, **CNN**, **and LSTM**. Maximum accuracy was obtained by CNN and LSTM with 86% and 85%, respectively, proving to be effective in Bangla text emotion detection. They collected 36,000 Facebook comments in Bangla related to the following six emotional states: joy, sorrow, rage, contempt, surprise, and terror. Indeed, it took a lot of preparation over the quality of the dataset, with tokenization and the removal of noisy information. The features were extracted using word embedding and TF-IDF, gathering statistical and semantic elements from the text. Data was drawn from a single source, Facebook, which may create biases and limited generalizability across Bangla text genres or domains. Though the preprocessing is reliable, feature representation can be further improved by using more sophisticated approaches in languages, such as stemming or domain-specific vocabulary.

Sadia Afrin Purba et el. [11] paper proposes a scheme for document-level emotion detection- "happy", "sad", and "angry " in Bangla text. In this paper, the authors evaluate a new dataset of 995 hand-annotated Bangla documents using ML and DL models: 995 Bangla texts in total divided into 249 for testing and 746 for training. The pre-processing steps include symbol removal, stemming, and stopword removal. For feature extraction, TF-IDF, word embeddings, and count occurrences were used. Some utilised models are CNN, ANN, Multinomial Naïve Bayes, and LR. Among them, Multinomial Naïve Bayes performed best with 68.27% accuracy.

A machine learning technique based approach for sentiment detection is proposed in [12], in this paper they have also analyzed some features but did not actually use them in their research. They performed a binary classification using tf-idf classifier to find out the most informative words and got 83% accuracy using this approach.

A Bangla tweet sentiment polarity detection using word and character n-grams along with Naïve Bayes has been proposed in [13]. Authors also looked at the Sent WordNet feature, which is a lexical resource of sentiment polarity analysis. They classified the tweets using Multinomial NB. Using 1000 training and 500 test data, they achieved 48.5% accuracy.

2

In another paper [14], a lexicon-based backtracking approach has been used over 301 test sentences for binary emotion classification. In this paper, they first classified the sentiment of the data and then the emotion. The dataset was mainly collected from Facebook status, news headlines, textbook, and direct speech. They claimed an accuracy of 77.16% using this approach.

A good study on emotion tagging has been done on paper [15]. In this paper, authors aimed at doing the manual annotations of sentence-level text from web-based Bengali blog corpus and observed the classification results on the corpus they annotated. With 1200 training instances, The Conditional Random Fields (CRF) classifier gave them an average accuracy score of 58.7% and Support Vector Machine (SVM) managed to get them to 70.4%.

53

In paper [16] proposes a machine learning-based approach for Bangla text emotion identification. The present study focuses on the much less explored area of emotion recognition for Bengali literature. It outlines a framework for classifying the six primary emotions of fear, surprise, anger, sadness, happiness, and disgust. addressed the dearth of resources in Bangla language processing by creating a special corpus with 1200 tagged samples of Bangla emotions. Techniques for preparing data, including feature extraction and tokenization, are well explained. makes use of feature extraction techniques that are appropriate for the task, such as Bag-of-Words and CountVectorizer. When compared to the Naive Bayes classifiers, SVM outperforms them with an accuracy of 73%.The dataset size is too short, which makes the model not robust and less generalizable (1200 text samples).The system only handles pre-labeled test samples. There is no mention of applying the method to the real world.

The summary of literature review is detailed in Table 2.2.1 below, which we utilized to guide our next steps.

TABLE 2.2.1 SUMMARY OF LITERATURE REVIEW

References	Year	Problem Dealt with	Size of image Data Set	No. of Classes	Algorithm	Findings
This Paper	Null	Emotion Detection from Bangla Text	4102	7	Logistic Regression, RNN, DNN, Multinomial Naive Bayes, Gradient Boosting Classifier, CNN, Bangla Bert, Random Forest Model, SVM, Random Forest Classifier	Best Algorithm Model: CNN Accuracy : 88.64%

S. Azmin and K. Dhar [3]	2019	Classify the text into three different classes Happy, Sad, Angry	420	3	Multinomial Naïve Bayes	Best Algorithm Model: Accuracy: 78.6%
Rowshan Rahma n Rushan et al. [4]	2024	structured approach to developing a system for detecting emotions in Bangla text and associating them with emojis	9940	3	Random Forest, SVM, Naive Bayes, SGD Classifier, BERT-based Deep Learning	Best Algorithm Model: Bert Based Deep Learning Accuracy: 85%
Abdullah Al Jamil ,Rifat Rahm an[5]	2021	Sentiment positive , negative and None, Disgust/Anger, Surprise/Fear	7732	5	BiLSTM, BiGRU, CNN, NB, SVM, RF, DT	Best Algorithm Model: CNN CNN Model Accuracy: 83%

Amit Kumar Das et.el. [6]	2020	Hate Speech Detection	7425	7	CNN, LSTM, GRU,	Best Algorithm Model: CNN+ Attention model Accuracy : 77%
Shahidul Islam Khan et. el.[7]	2022	Detecting Anger, Fear, Joy, Sadness	600	4	Logistic Regression, Naive Bayes, LSVM, Random Forest	Best Algorithm Model: LSVM Model Accuracy : 74.3%
Moshieur Rahaman Faisal et el. [8]	2024	Detecting emotion from Bangla and Banglisch Text	80098	6	Random Forest, Naïve Bayes, CNN, BiLSTM, BanglaBert , XML Roberta	Best Algorithm Model: Bangla Bert Model Accuracy : 71.3%

Md Ataur Rahman , Md Hanif Seddiqi[9]	66 2019	Comparison of classical machine learning approaches on Bangla textual emotion analysis	6314	6	① KNN, Naïve Bayes, Decision Tree, Support Vector Machine, K-means clustering	Best Algorithm Model: SVM Model Accuracy : 52.98%
Tahmina Akter et. el. [10]	2024	⑫ Evaluating Machine Learning methods for Bangla text analysis	36000	6	⑫ KNN, Decision Tree, Random Forest, SVM, CNN, LSTM	Best Algorithm Model: CNN Model Accuracy : 86%
Sadia Afrin Purba et al. [11]	58 2021	Document Level Emotion Detection from Bangla Text Using Machine Learning Techniques	995	3	Logistic Regression, Multinomial naive bayes, Artificial Neural Netwrok, Convolutional Neural Network	Best Algorithm Model: Multinomial Naïve Bayes Accuracy : 68.27%

## 2.2.1 Related Research

### 2.3 Gap Analysis

As we can see that there is no study found that has worked on seven classes. Work on sentiment analysis has been done on a vast amount but emotion detection hasn't been done that much. Additionally, the data collection was from single source either

YouTube or Facebook or Twitter. But the users are different from every site. So, this is also a gap since age and environment and platform varies a lot and the expression of emotion is also very different.

## 2.4 Summary

This work uses CNN-based models to classify different kinds of nuts to fill a gap in <sup>1</sup> the current state of the art in research conducted on the topic. Particularly in developing nations, nuts are an important agricultural product with the potential to help reduce the hunger crisis. Although previous studies proved that the deep learning (DL) methods are effective in agriculture applications, such as disease detection and crops classification, there are much more limited investigations on the nut breeds recognition. In this study, CNN-based models are examined to efficiently and accurately classify six types of nuts. These findings could have several applications including enhancing agricultural productivity, accuracy and boosting developing countries like Bangladesh economically.

## **RESEARCH METHODOLOGY**

This chapter highlights the research design, methods of data collection, and experimental method adopted for the study. It outlines the methodology and instruments used to achieve accurate identification

### **3.1 Methodology**

It articulates the structured approach for research, involving preparation strategies, model selection, training, and assessment. The section provides a systematic record of methodologies adopted to identify breeds of nuts accurately.

#### **3.1.1 Overview**

The methodological section of our study tries to present to the readers in detail all the methods and techniques that have been followed in carrying out the inquiry. In this part of our study, we go into detailed descriptions of the methods that were employed to generate, preprocess, and refine the data that we have in hand. We also provide illustrations of the gathered data and associated procedures. We also discuss the classification problem we want to solve with our proposed model, the specific classification models used, the different levels and their components in these models, and the data set assembled to achieve the best accuracy and minimize the loss. This section presents a detailed and profound analysis of the methods used for information production, preliminary processing, enhancement, graphical representation, and choice and building of categorization frameworks. In such a way, it is possible to solve the current research problem by creating a valid and relevant approach. It was really significant for us to follow the data gathering and data cleansing steps because the process depended on them. In training the classification model, we also faced some challenges as the best strategy for training was tricky to choose.

### 3.1.2 Proposed Methodology

This is the Methodology we have come up with:

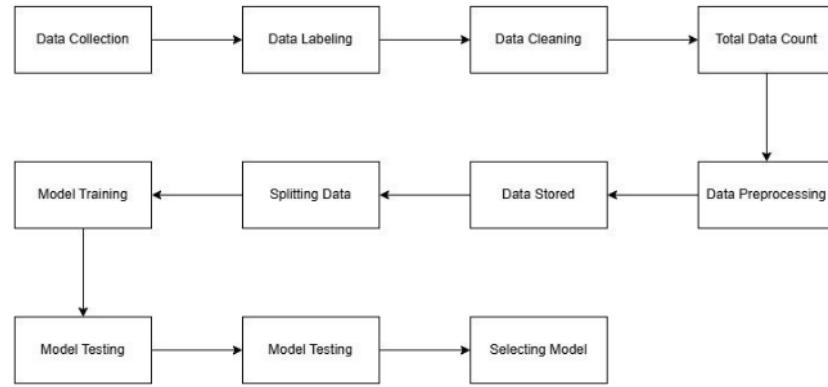


Figure 3.1.2.1 Methodology Process

## 3.2 Detailed Methodology and Design

1

### 3.2.1 Data collection procedure

The dataset collection process for the Bangla text was, therefore, not an end in itself but a means to ensure that the proposed system was robust and reliable. This section describes the step-by-step process adopted to collect, select, and annotate the data required for the research into multilabel emotion recognition in Bangla text.

The text data in Bangla was collected from a wide range of publically accessible sites to ensure that different varieties of language styles, emotional expressions, and settings are covered. Some of the primary sources were:

Social Media: Now a days people convey their emotion more in social media.

Data was collected from Facebook, YouTube.

E-commerce website: Data was collected from customer review given on the products they purchased and used. Daraz, Shajgoj, Chardike.

TABLE 3.2.1.1 SAMPLE COLLECTED DATASET

Sentence	Label
ନବବଧୂର ଜଳ୍ୟ ଏର ଚେଯେ ମେରା ଉପହାର ଏକଜଳ କ୍ରିକେଟାର ସ୍ଥାମୀର ଜଳ୍ୟ ଆର କି ହତେ ପାରେ କୁଣ୍ଡ କୁଣ୍ଡ କୁଣ୍ଡ	Happiness
ବୁଦ୍ଧି ଜ୍ଞାନ ଏକେବାରେ ଶୁଣ୍ୟେର କୋଠାୟ	Sadness
ଓଦେରଓ ବିଷ ଥାଓଯେ ମାର	Angry
ଏଗ୍ରଲାର ଦରକାର କି ?	Confused
ରାଷ୍ଟ୍ରା ଦିଯେ ହେଟେ ଯାଓଯାର ସମୟ ହଠାତ୍ କୁକୁର ଦେଖେ ପିଛିଯେ ପଡ଼ିଲାମ।	Fear
ଏଟା ଭାବଲେଇ ଗା ଘିନଘିନ କରଛେ।	Disgust
ବିଶ୍ୱାସଇ କରନ୍ତେ ପାରଛି ନା, ଏଟାଇ ସତିୟ!	Surprise

### 3.2.2 Dataset Cleaning

First, we imported the dataset into the 'Google Colab'. Then we extracted some information like how many unique data, how many null values are there. Then it was checked if there is any spelling mistake. After that, we dropped the null value, and spelling mistakes were corrected. After that, we drew a graph.

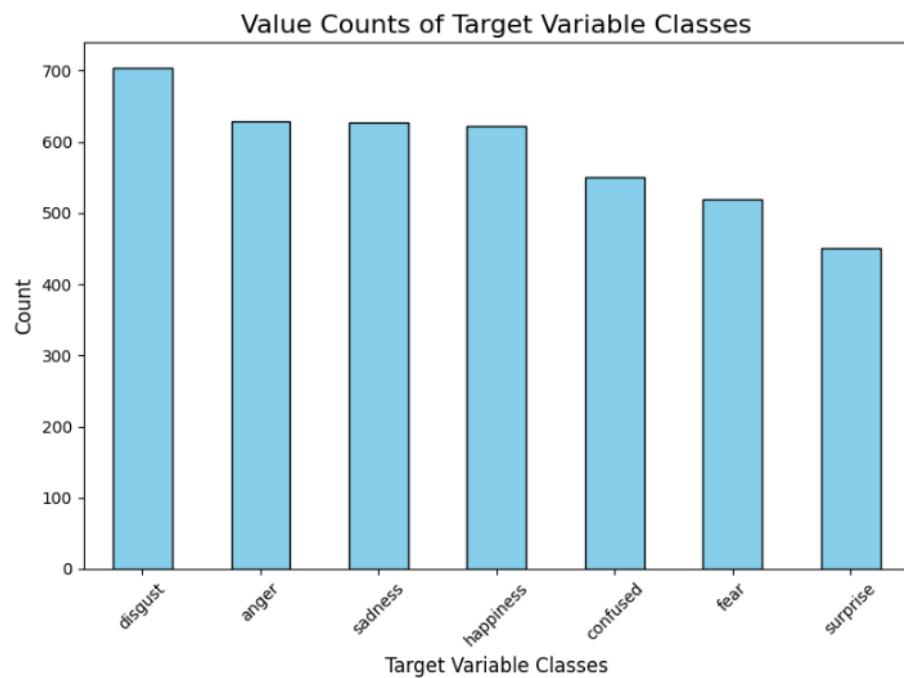


FIGURE 3.2..2.1 DATA COUNT OF SEVEN CLASSES

Table 3.2.2.1 The Final Dataset Table

EMOTION	COUNT
DISGUST	704
ANGER	629
SADNESS	627
HAPPINESS	622
CONFUSED	550
FEAR	520
SURPRISE	450

### 3.2.3 Dataset Preprocessing

Data preparation is an important phase of the data assessment segment. It involves <sup>83</sup> cleaning, converting, and arranging raw data in a manner to make it ready for evaluation and training models. In Natural Language Processing, the most important part is data cleaning, as the data is not in a shape from which it can be used to train. There is a long process to follow for the data preprocessing step.

**Removing Emoji:** Pattern recognition was used to remove the emoji at the very first. But after applying Pattern Recognition there were a few emoji left. Then emoji library was used to remove remaining emoji. That's how all the emoji was removed from the cleaned dataset.

Table 3..2.3.1 Cleaned Dataset Table

Sentence	emotion	n_emotio n	Cleaned sentence
অনেক হইছে ভাই। এইবার শাটার নামান। 😊	anger	1.0	অনেক হইছে ভাই। এইবার শাটার নামান।
অনেক মজা পাইসি ভাই 😊	Happines s	2.0	অনেক মজা পাইসি ভাই
ওই লোকটা মারা গেসে 😭	Sadness	5.0	ওই লোকটা মারা গেসে

<sup>24</sup>

**Tokenization:** Tokenization is the process of breaking down text into smaller units called tokens. These tokens can be words, characters, or even subword units. This process is a basic step <sup>67</sup> for many NLP tasks, such as sentiment analysis, machine translation, and text summarization. From bnlp we used Basic Tokenizer to tokenize. Some of the sentences were not tokenized. Then dataset was tokenized again using NLKT.

Table 3.2.3.2 Tokenized Dataset Table

Cleaned Sentence	Tokenized
হায় হায়, এটা কী হলো!	[হায়, হায়, এটা, কী, হলো, !]
ওরে না, এমন কিছু আগে কখনো শুনিনি!	[ওরে, না, এমন, কিছু, আগে, কখনো, শুনিনি, !]
পায়ের উপর পা দিয়ে বসাটা কি ভদ্রতা?	[পায়ের, উপর, পা, দিয়ে, বসাটা, কি, ভদ্রতা, ?]

**Stop word, Punctuation, Vowel, Digits, Single Letter removing:**

**corpus.stopwords:** A collection of frequently used Bengali stopwords , such as 'অতএব', 'আথচ', 'আথবা', 'অনুযায়ী', 'অনেক', 'অনেকেই', 'অন্তত' and so forth, that are typically omitted from text processing tasks.

**Corpus Punctuations:** A list of Bengali punctuation symbols. !"#\$%&'()\*+-./:;<=>?@[|^`\_{|}~।ংঃঁ

**Bengali letters:** All Bengali letters are listed in corpus letters.

অআইউউখএঞ্চওঁকখগঘঙ্গচছজঘঞ্চটঠডণতথদধনপফবভময়ৱলশষসহড়ঢ়য়ঃঁ

**Bengali Digits:** A list of Bengali digits is called corpus digits. ০১২৩৪৫৬৭৮৯

**Corpus vowels:** A list of vowel characters in Bengali. ঔ ি ি ি ি ি ি

**English letter, punctuation, digits remove:** After removing Bengali digits, punctuations, vowel, English letters, digits were removed.

**Dropping null value:** After taking these steps, some null value has been found. Then these null values have been removed.

**Cross word validation:** The validation of the words is needed since a lot of preprocessing has been done on the dataset. To check if the words are in the Bangla dictionary or not. <https://github.com/tahmid02016/bangla-wordlist?tab=readme-ov-file#basic-tokenizer> from this link we have validated the preprocessed words. Currently it contains 454650 words and it is sorted. Till now preprocessed words are checked and validation was ensured.

**Lemmatization:** After doing lemmatization, a few words were found that are not in the dictionary. “লাগে” is a Bangla word. After lemmatizing this word, it becomes “লা.” But

there is no “না” in that word list. So, we didn't change it. “বৎসরে” and “বঞ্চিত” couldn't be lemmatized. So, it remained the same.

**Stemming:** After lemmatization we used Bangla stemmer for stemming. Then the word “৩০০০” was turned into “০০০”. We dropped the null values after the stemming step.

**Feature Extraction:** One of the efficient and popular ways for feature extraction in applications regarding emotion detection is TF-IDF vectorization. It means converting unstructured text into structured numerical data, which will be good to go for the machine learning algorithms.

### Term Frequency

The frequency of a word within a document is the number of times it occurs. It gives an idea of the meaning of the word in a local sense in a given document.

$$TF(t) = \frac{\text{Total number of terms in a document}}{\text{Number of times term } t \text{ appears in a document}}$$

This makes it easier to find terms that appear frequently in a document and gives an indication of how relevant they are to that particular context.

### IDF is the frequency of inverse documents:

Considering the word frequency across the entire set of documents of the corpus, the inverse document frequency calculates how much information a word conveys. Common stopwords and other words frequent in many papers are given lower IDF, while words occurring seldom in many documents are given higher IDF, signifying the higher importance or distinctiveness in the document.

$$IDF(t) = \log\left(\frac{\text{Total number documents}}{\text{Number of documents containing term } t}\right)$$

Formula:

$$TF-IDF(t,d) = TF(t,d) * IDF(t)$$

t is the term

d is the document

84 **Data splitting for training and testing:** We split the data into two parts for machine learning. They are training and testing. Eighty percent data is used for training and twenty percent data is used for testing. For Deep learning twenty percent data is used for

#### 54 3.2.4 Logistic Regression

Logistic regression is one of the most popular statistical and machine learning approaches for classification tasks. Despite its name, it is actually a method for binary or multiclass classification rather than a regression method. It predicts the probability that a data point belongs to a certain class by using a logistic (sigmoid) function to model the relationship between the input attributes and the target class.

Logistic Regression model is initiated with a fixed random state and a maximum of 1000 iterations.

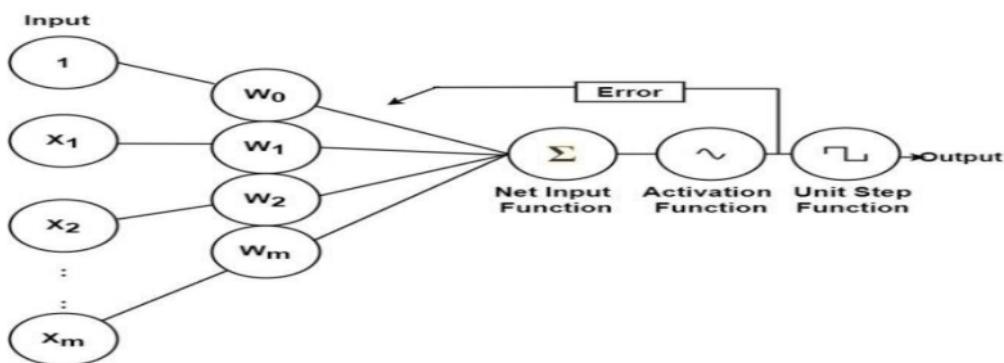
Then, the `fit()` method is used to train it on `X_train` and `y_train`.

The predictions are then stored in `y_pred_log_reg` after the model uses the `predict()` method to generate predictions for test data that has not been seen, `X_test`.

If  $p$  is the probability that a binary response variable  $Y = 1$  when input variable  $X = x$ , then the logistic response function is modeled as:

$$p = P(Y = 1|X = x) = \frac{e^{\beta_1 + \beta_0 x}}{1 + e^{\beta_1 + \beta_0 x}}$$

This function non-linear and **s** shaped curved [17] In this training we have set hundred iterations and random state is forty two.



97 Figure 3.2.4.1 Logistic Regression Architecture

### 3.2.5 Random Forest Classifier

17

The Random Forest Classifier is a robust and flexible machine learning algorithm that belongs to the family of ensemble learning techniques. It works particularly well on applications involving both regression and classification. During training, the Random Forest algorithm constructs multiple decision trees and returns a class that is the mean prediction (regression) or the mode of the classes (classification) of the individual trees.

69

N estimators=100 is the parameter informing about the number of Decision Trees to be constructed in the forest. It is set to 100 in this example, although the larger the value the better the model's accuracy, and it is at the cost of the computing time and other resources.

6

Random state=42: This ensures reproducibility in the results. By setting a random seed, we ensure that the same split and the same random model selections are returned every time we run the code.

6

6

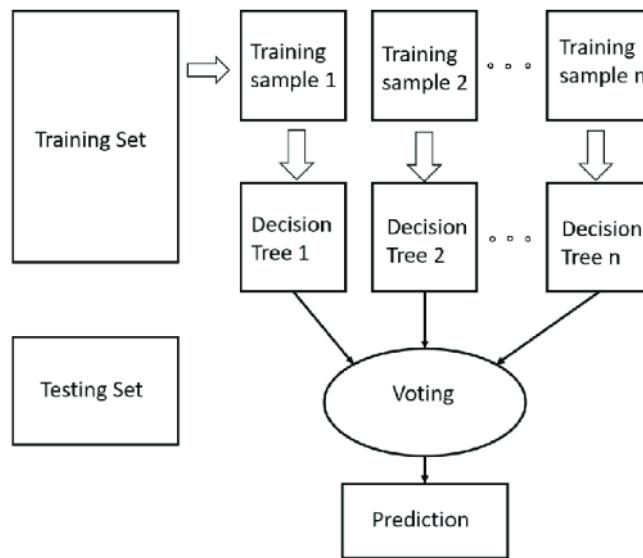


Figure 3.2.5.1 Random Forest Classifier Architecture

32

### 3.2.6 Multinomial Naive Bayes Classifier Model

Naïve Bayes is a simple and efficient probabilistic machine learning technique based on Bayes Theorem. It is generally used for classification problems and is particularly popular in NLP for tasks like text classification, sentiment analysis, and spam detection.

X\_train: Feature matrix of training data, generally preprocessed text data, for instance word frequency counts or TF-IDF.

y\_train: Target labels that match the training data.

17

The algorithm now calculates the prior probability of each class.

$$P(Y = c) = \frac{\text{Number of samples in class } c}{\text{Total number of samples}}$$

For a particular class, it then determines a feature's probability:

$$P(x_i|Y = c) = \frac{\text{Count of feature } x_i \text{ in class } c + \alpha}{\text{Total count of all features in class } c + \alpha n}$$

The Laplace smoothing parameter prevents zero probabilities to unseen features.

38

### 3.2.7 Support Vector Machine

The Support Vector Machine is a popular supervising machine learning approach for classification and regression problems. It performs very well in applications related to image recognition, sentiment analysis, and text classification, especially in high-dimensional areas. SVC: This is the Scikit-learn implementation of the Support Vector Classifier for classification problems. kernel='linear': This suggests that a linear kernel is used by the SVM. The type of decision boundary is determined by a kernel function. random\_state = 42. This sets the seed to ensure that any random processes in the model (if any) are reproducible. X\_train: Feature matrix of the training set. y\_train: Target labels of the training set. The algorithm SVM determines the best hyperplane dividing the data points of various classes with the maximum margin at train time.

57

### 3.2.8 Gradient Boosting Classifier

The Gradient Boosting Classifier is an ensemble learning algorithm useful in classification problems. This class of methods leverages multiple weak learners, typically decision trees, to produce a strong predictive model by combining them sequentially. The approach uses gradient descent in a step-by-step minimization of the prediction error to optimize for accuracy.

In classification problems, GradientBoostingClassifier initiates a Gradient Boosting model. This is an ensemble method in which several decision trees are sequentially added to improve the performance. random\_state=42: The random seed is fixed for reproducibility. This ensures that running the same code multiple times will have the same result.

7 fit(X\_train, y\_train): Fit the Gradient Boosting Classifier according to the training data.

X\_train: The training input feature matrix.

6 y\_train: Target labels for training data.

6 predict(X\_test): Predicting the class labels for test dataset using the trained model of Gradient Boosting, where X\_test is called the input feature matrix for invisible data.

### 3.2.9 CNN

26

One variant of deep learning, the Convolutional Neural Network or CNN, is designed primarily to analyze data that comes in a structured grid format such as time-series data, images, etc. They are particularly suited to tasks like speech recognition, object detection, picture classification, and natural language processing.

In NLP tasks, the length of the text sequences is often different. Longer sequences are cut down to a certain size, and shorter ones are padded with zeros since neural networks require input sequences of the same size.

Pad sequences: Pads a sequence to a maximum length (maxlen) that is supplied. We have taken the maxlen=100. Ensures that each sequence exactly has 100 tokens. We have taken 128 layers in each convolutional layer. We have taken embedding vector size as 100. To introduce non-linearity we have used Relu activation. To prevent overfitting we have randomly dropped fifty percent of the Neurons.

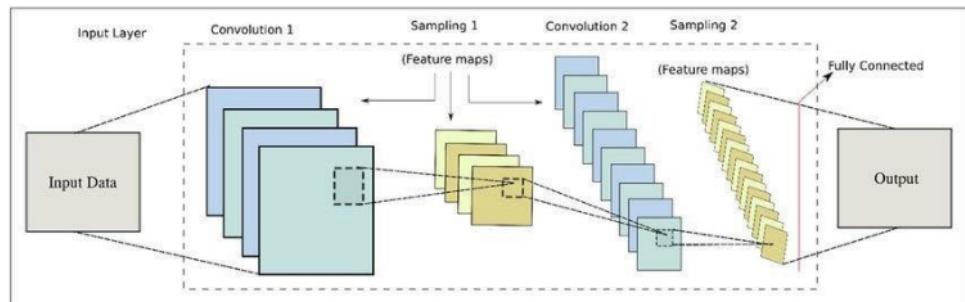


Figure 3.2.9.1 CNN Architecture

### 3.2.10 Bangla Bert

Bangla BERT is a transformer-based language model that has been specially pre-trained on Bengali (Bangla). Based on the popular BERT concept—which was first presented by Google<sup>96</sup> Bidirectional Encoder Representations from Transformers, it is built to handle the linguistic subtlety, vocabulary, and syntax of Bangla.

Exchanges Bangla text in the custom dataset BanglaDataset using the Bangla BERT tokenizer (sagorsarker/bangla-bert-base). Sequences will then be padded or truncated up to a specified length before transformation into PyTorch tensor blocks (max\_len= 128). PyTorch DataLoader has been used for batching with shuffling for training and validation datasets. Instantiating and setting up a pre-trained Bangla BERT model for the purposes of sequence classification with seven labels.

The training loop keeps track of progress using tqdm and adjusts weights using the AdamW optimizer. Pre-trained weights are adjusted to fit the particular dataset as the model is fine-tuned over five epochs.

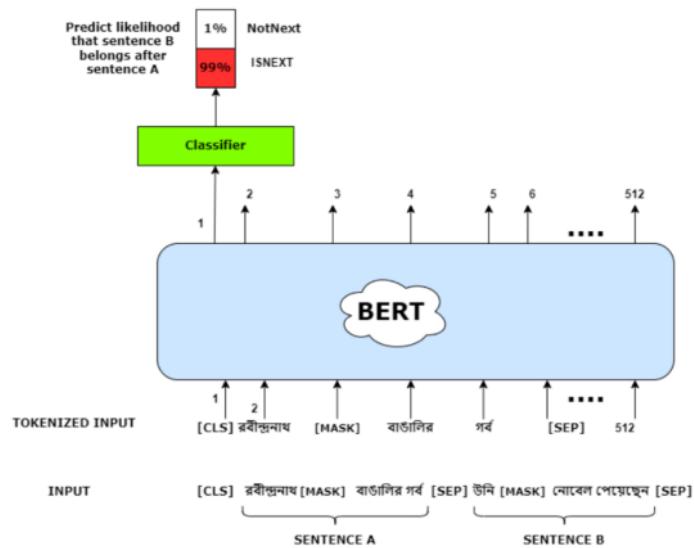


Figure 3.2.10.1 Bangla Bert model Architecture

### 3.2.11 RNN

One kind of artificial neural network made for sequential data, including time series, text, audio,

or video, is called a recurrent neural network (RNN). By preserving a hidden state that records details about prior inputs, RNNs are able to process sequences of different lengths, in contrast to conventional feedforward neural networks.

We used Tokenizer to convert the preprocessed Bangla text data into sequences of numbers. For homogeneity, pads sequences to a pre-specified length: max sequence length=100. We used LabelEncoder to encode categorical target labels (like emotion) into numerical representation. It captures meaningful links by mapping words to dense vectors (embedding\_dim=100).

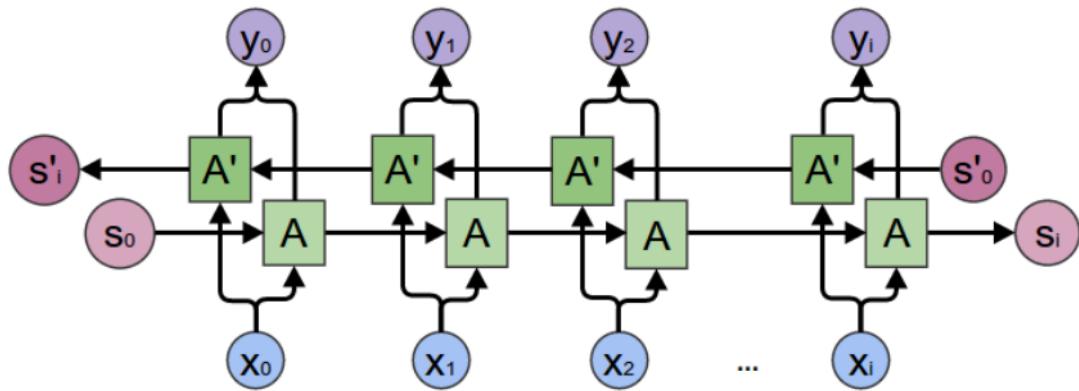


Figure 3.2.11.1 RNN architecture

### 3.2.12 DNN 72

A deep neural network (DNN) is a type of ANN that contains more than one layer between the input and output layers. Having a large number of hidden layers enables the network to learn hierarchical patterns from input, thus the name "deep" in this context. DNNs are widely used nowadays in such areas as computer vision, natural language processing, speech recognition, and many more. They form the basis for many complex machine learning tasks.

We used a tokenizer to convert the Bangla language into number sequences. Pad\_sequences ensures that the input will have a fixed length, here max\_sequence\_length=100. We used LabelEncoder to convert categorical labels (emotion) into numeric representation. We split the

dataset into 20% test sets and 80% training sets. In semantic representation, the word indices are changed into dense word embeddings with size embedding\_dim=100. ReLU activated dense layers to learn and extract features. Dropout Layers <sup>86</sup> go ahead and randomly turn off neurons while training to avoid overfitting. trains the model, with a batch size of 128, in 10 epochs, monitoring loss and validation accuracy.

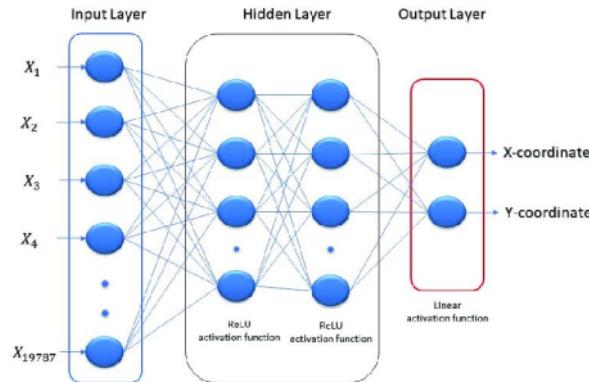


Figure 3.2.12.1 DNN model architecture

### 3.3 Project Plan

The project plan delineates the organized timetable, resources, and milestones necessary for the effective implementation of this research on nut breed recognition with deep learning. It offers a framework to guarantee the methodical advancement of the research while conforming to established objectives and timelines.

Table 3.3.1 Project Plan

Phase	Activities	Duration	Outcome
Phase 1: Problem Identification and Literature Review	What's the problem and how should I solve this. The previous work or paper had been searched or read.	2 weeks	I gained knowledge about the previous paper and had an idea how to proceed in future

Phase:2 Data Collection and Preprocessin g	Collecting the data from Social Media, E-Commerce sites and annotation them	4 weeks	A diverse dataset was created
Phase 3: Model Development	We developed nine models	4 weeks	Optimized models for detecting emotion from Bangla Text
Phase 4: Model Evaluation andAnalysis	We have evaluate the model And was testing	2 weeks	The performance done by the models
Phase 5: Result Visualization and Documentation	Models result visualization	1 week	Confusion Matrix, Training and Validation accuracy and training and validation loss
Phase 6: Finalization and Submission	Revise the paper again and try to find out any mistake	1 week	Submitted the paper to our supervisor

The 3.3.1 project plan table gives us an Idea about the time schedule and step by step proceeding towards the goal

### **3.4 Task Allocation**

The thesis had been done only by me. So, I had to all the task all by myself.

### **3.5 Summary**

In this thesis, several machine learning and deep learning techniques were employed for emotion categorization in the Bangla language. It uses some conventional models like Random Forest, Support Vector Machine, Logistic Regression, Gradient Boosting Classifier, and Multinomial Naive Bayes for a baseline comparison. The complex patterns of text were captured using the sophisticated design of CNN, RNN, and DNN. A pre-trained transformer-based model, Bangla BERT, was fine-tuned to enhance the contextual understanding of the Bangla language.

## CHAPTER 4

### Implementation and Results

#### 4.1 Environment Setup

##### Programming Language and Frameworks:

**Python** : Selected due to the availability of a rich library and framework for  
machine learning and natural language processing.<sup>27</sup>

Key libraries:

**Scikit-learn**: Traditional machine learning models such as SVM, Random Forest,  
Logistic Regression, and Naive Bayes.<sup>51</sup>

**TensorFlow/Keras and PyTorch**: Deep learning models such as CNN, RNN, DNN,  
and Bangla BERT.<sup>27</sup>

**Jupyter Notebook or Google Colab**: For the development and interactive  
experimentation with the model.

**Integrated Development Environment (IDE)**: Used Visual Studio Code for  
maintaining project files and debugging.

#### Hardware

RAM:

At least 16GB is recommended to handle the dataset and model training.

#### Libraries for NLP:

NLTK, spaCy, and Bangla FastText for preprocessing and embeddings.

Data handling:

Pandas and NumPy for efficient data manipulation and analysis.<sup>56</sup>

Visualization:

Matplotlib and Seaborn for plotting performance metrics analysis results.

### **Dataset Storage:**

The dataset was stored in CSV format, and tools like Google Drive or local disk were used for storage during experimentation. This ensured a well-configured environment in which the implementation of machine learning and deep learning models would be easier to train and evaluate.<sup>85</sup>

## **4.2 Performance Analysis**

The performance analysis is the most crucial part since it will say which model works the best. We have calculate accuracy for each model we applied. We also evaluate the model using F1 score, Precision, Recall. We also determine the confusion matrix.

### **4.2.1 Accuracy**

Accuracy is the ratio of all correctly predicted cases to all cases.<sup>65</sup>

$$\text{Accuracy} = (TP + TN) / (TP + FP + FN + TN) \quad (1)$$

### **4.2.2 Precision**

Precision is the ratio of true positive predictions to all positive predictions, including both false positives and true positives.<sup>44</sup>

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

### **4.2.3 Recall**

Recall is the ratio of true positive predictions to the total number of actual positive occurrences, including both true positives and false negatives.<sup>5</sup>

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

(3)

#### **4.2.4 F1 Score**

F1 Score balances the trade-off between precision and recall by considering their harmonic mean.

47

$$F1\text{-Score} = 2 \times (Recall \times Precision) / (Recall + Precision)$$

(4)

34

#### **4.2.5 Confusion Matrix**

The confusion matrix is a table that summarizes the performance of a classification algorithm. It displays the counts of TP, FP, TN, and FN. The CM helps visualize how well the model is distinguishing between classes and it gives insights into specific areas where the model might be making errors.

### 4.3 Results and Discussion

We applied both deep learning and machine learning model to detect emotion from Bangla text.

A table is given that is showing accuracy, precision, recall, F1 score of every model that applied.

TABLE 4.3.1 Performance Analysis Metrics

Model	Types	Accuracy	Precision	Recall	F1 Score
Logistic Regression Model	Macro	69	70	69	68
	Avg				
	Weighted				
	Avg				
Random Forest Classifier		73	74	73	72
	Macro				
	Avg				
	Weighted				
Multinomial Naïve Bayes		66	70	66	66
	Macro				
	Avg				
	Weighted				
SVM		71	72	71	71
	Macro				
	Avg				
	Weighted				
	Avg				

<b>Gradient Boosting Classifier</b>	<b>Macro</b>	<b>69</b>	<b>71</b>	<b>69</b>	<b>69</b>
	<b>Avg</b>				
	<b>Weighted</b>				
	<b>Avg</b>				
<b>CNN</b>	<b>Macro</b>	<b>89</b>	<b>88</b>	<b>89</b>	<b>88</b>
	<b>Avg</b>				
	<b>Weighted</b>				
	<b>Avg</b>				
<b>Bangla Bert</b>	<b>Macro</b>	<b>86</b>	<b>86</b>	<b>86</b>	<b>85</b>
	<b>Avg</b>				
	<b>Weighted</b>				
	<b>Avg</b>				
<b>RNN</b>	<b>Macro</b>	<b>88</b>	<b>88</b>	<b>88</b>	<b>88</b>
	<b>Avg</b>				
	<b>Weighted</b>				
	<b>Avg</b>				
<b>DNN</b>	<b>Macro</b>	<b>87</b>	<b>88</b>	<b>87</b>	<b>87</b>
	<b>Avg</b>				
	<b>Weighted</b>				
	<b>Avg</b>				

81

We will be visualizing the machine learning model and deep learning model results.

### 1 4.3.1 Logistic Regression

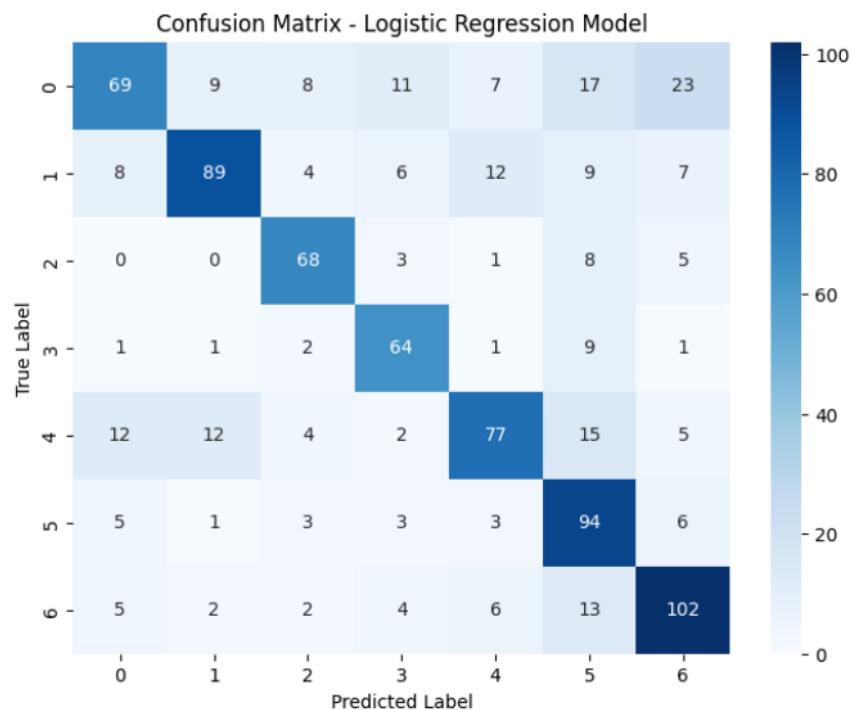


Figure 4.3.1.1 Confusion Matrix Logistic Regression

### 4.3.2 Random Forest Classifier

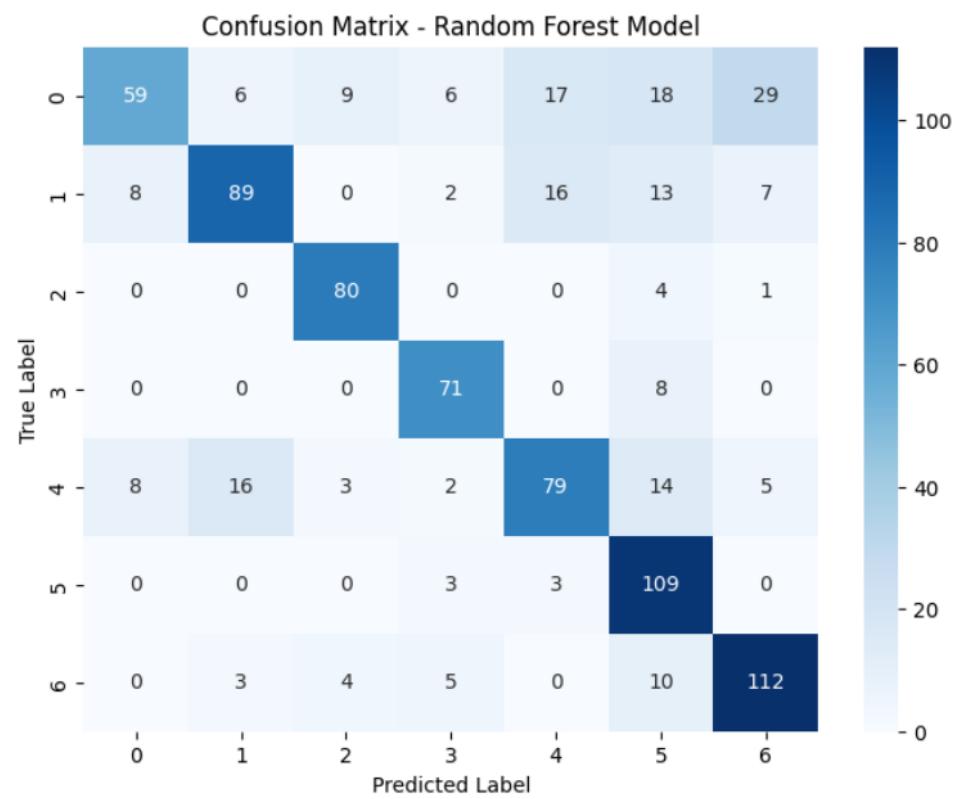
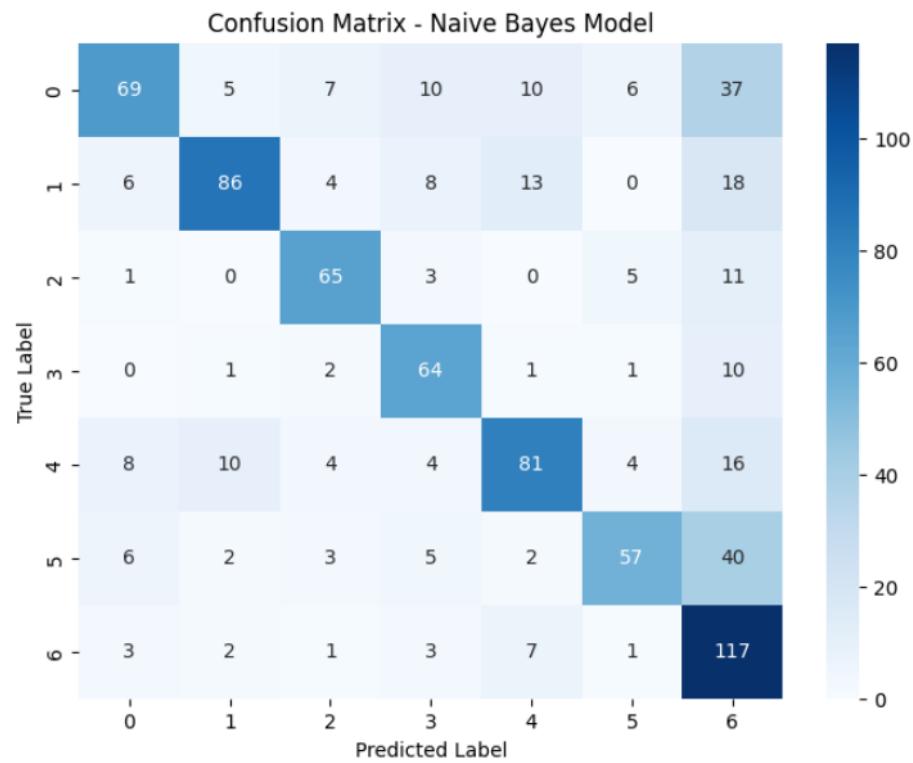


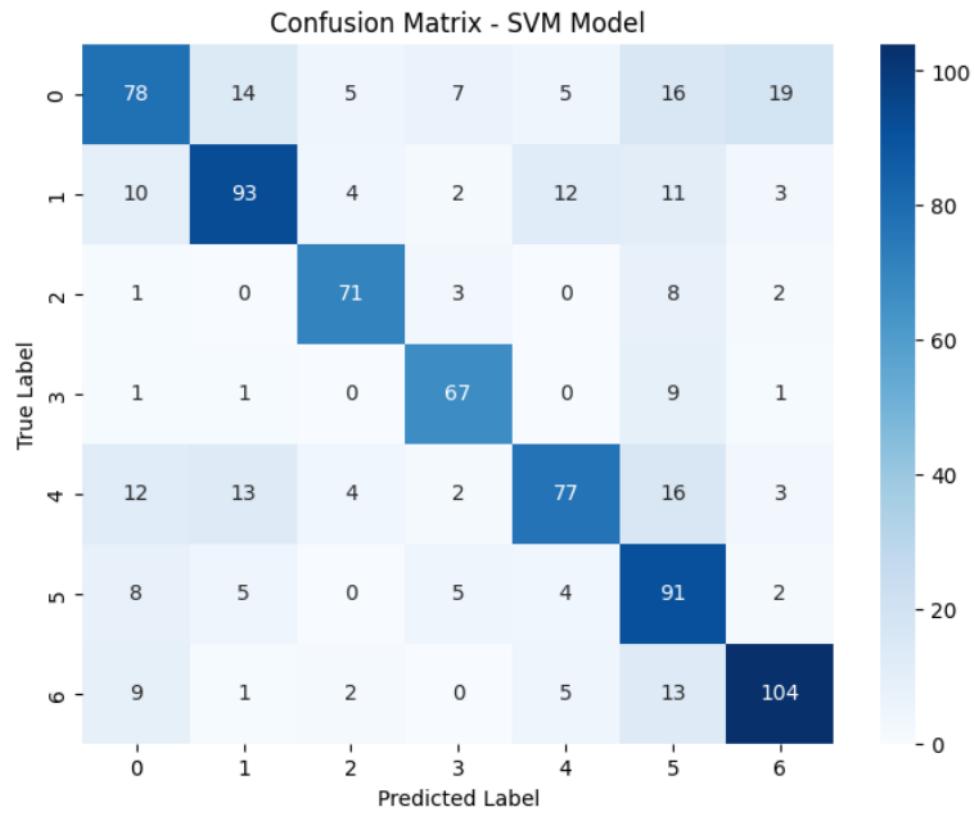
Figure 4.3.2.1 Confusion Matrix Random Forest Classifier

### 4.3.3 Multinomial Naïve bayes



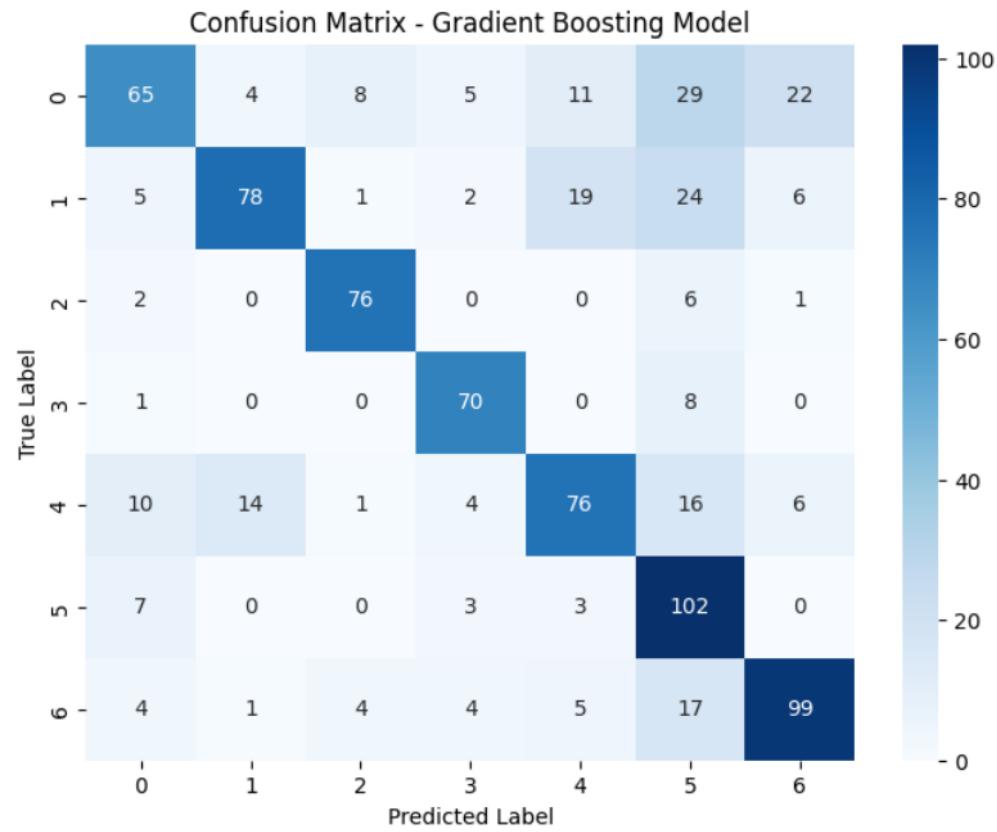
18  
Figure 4.3.3.1 Confusion Matrix Multinomial Naïve bayes

#### 4.3.4 SVM Model



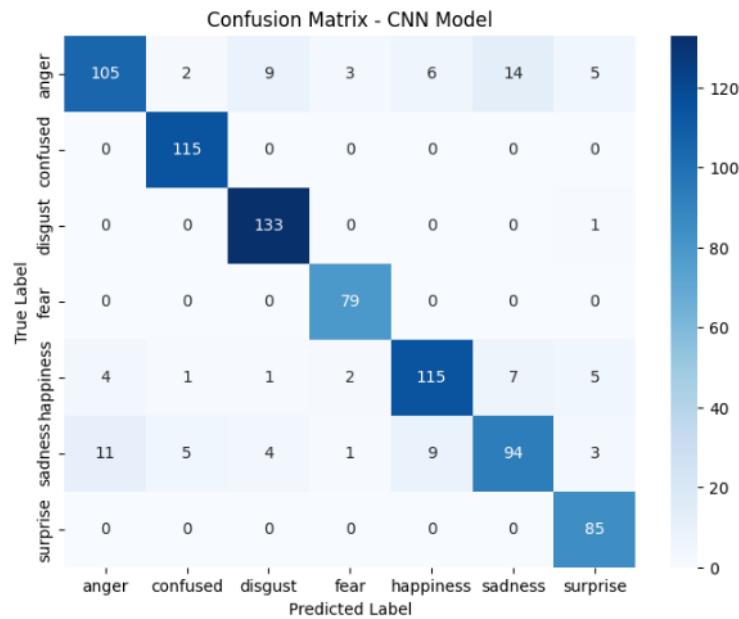
18  
Figure 4.3.4.1 Confusion Matrix SVM Model

#### 4.3.5 Gradient Boosting Model



18  
Figure 4.3.5.1 Confusion Matrix Gradient Boosting Model

#### 4.3.6 CNN



1 Figure 4.3.6.1 Confusion Matrix CNN Model

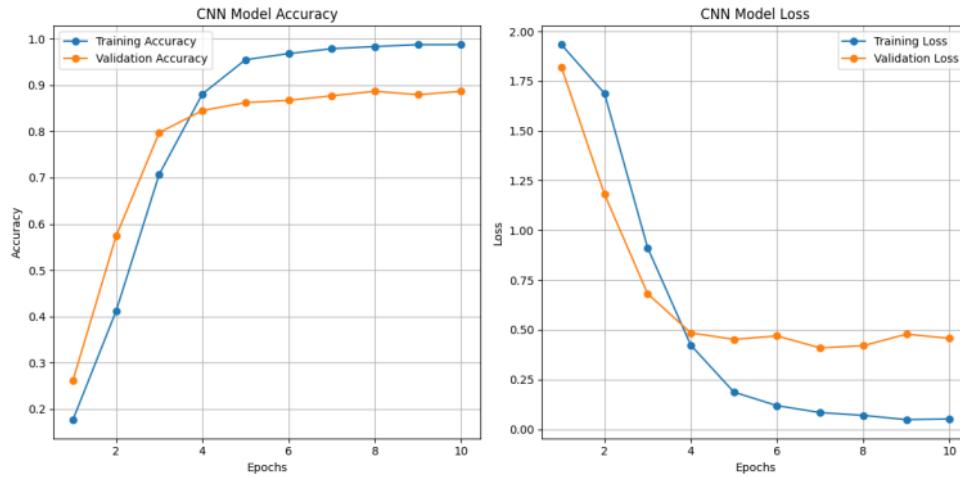


Figure 4.3.6.2 CNN Training and validation Accuracy

8 Figure 4.3.6.3 CNN Model Training and validation loss

#### 4.3.7 Bangla Bert

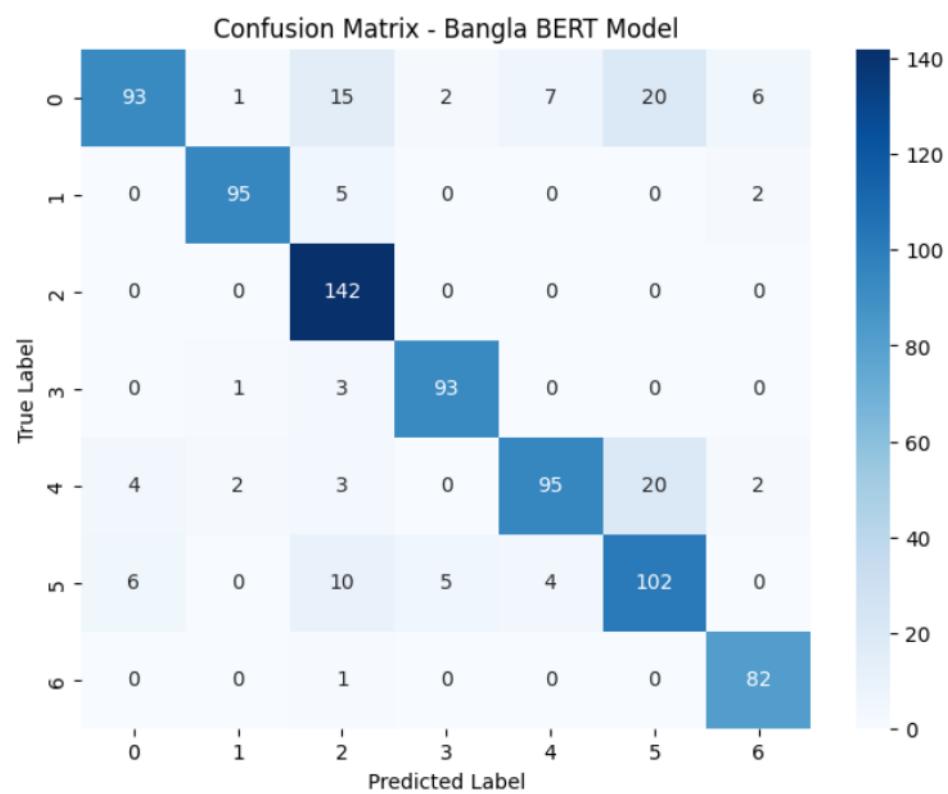


Figure 4.3.7.1 Bangla Bert Confusion Matrix

### 4.3.8 RNN Model

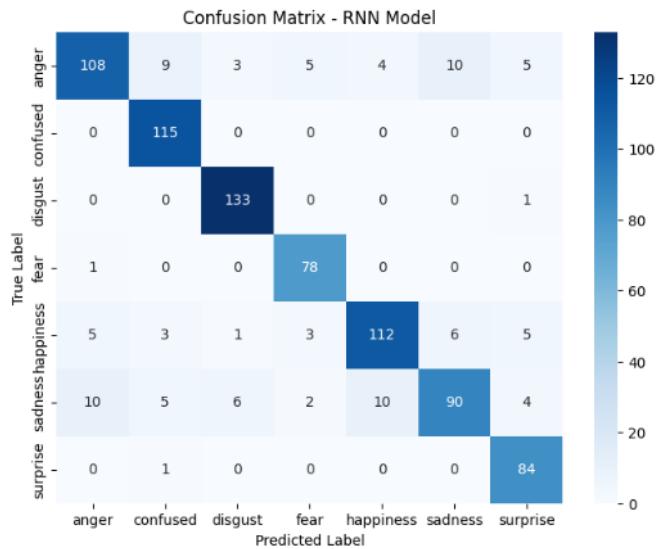


Figure 4.3.8.1 RNN confusion matrix

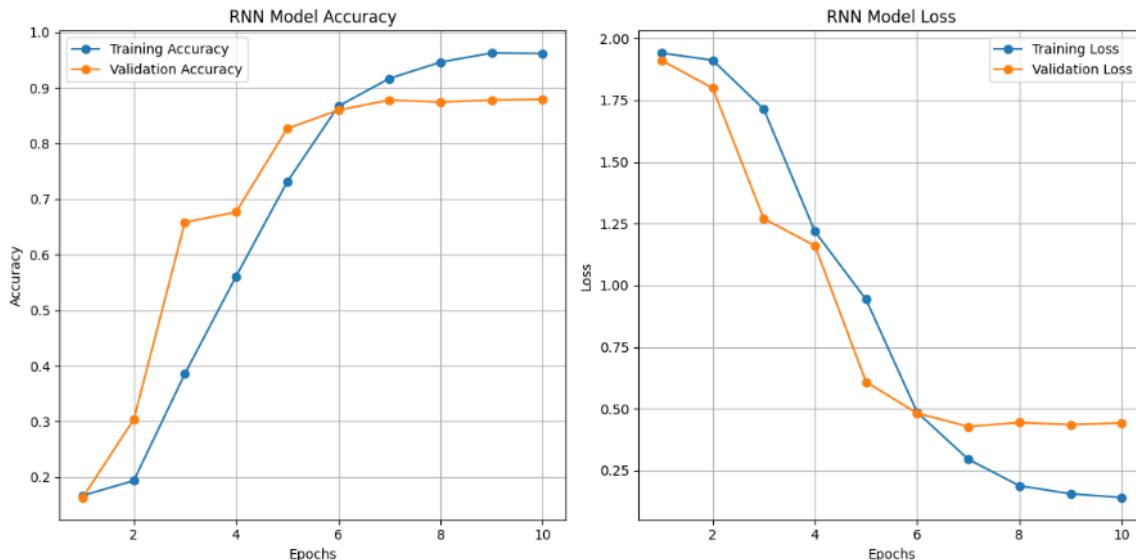
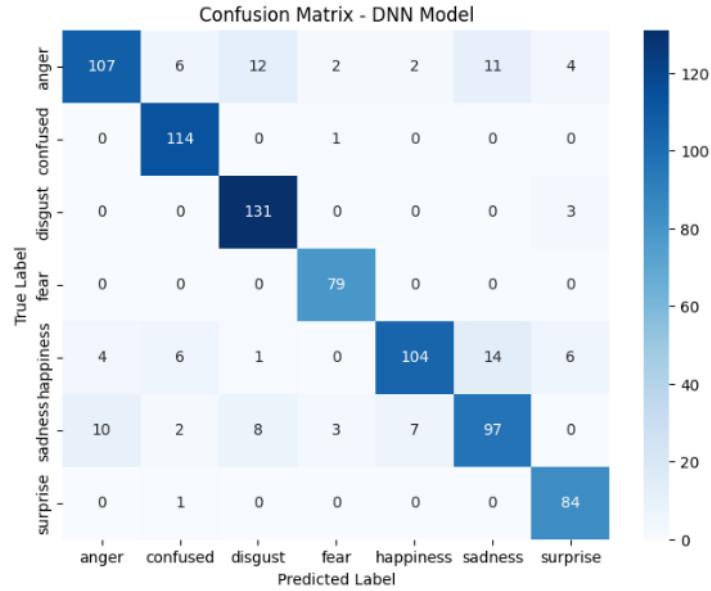


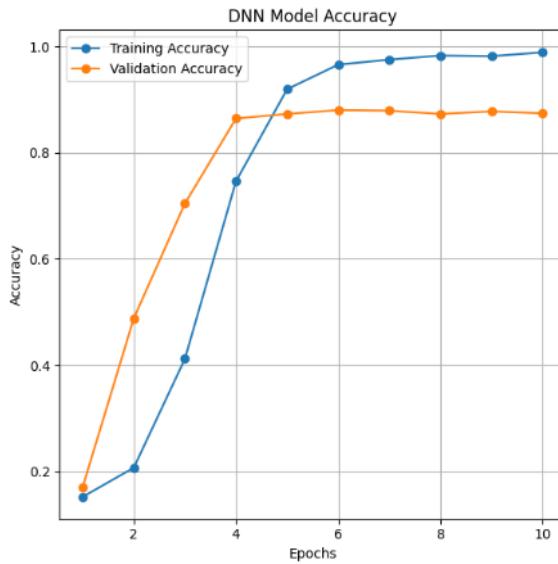
Figure 4.3.8.2 RNN Training and Validation Accuracy

Figure 4.3.8.3 RNN Training and Validation Loss

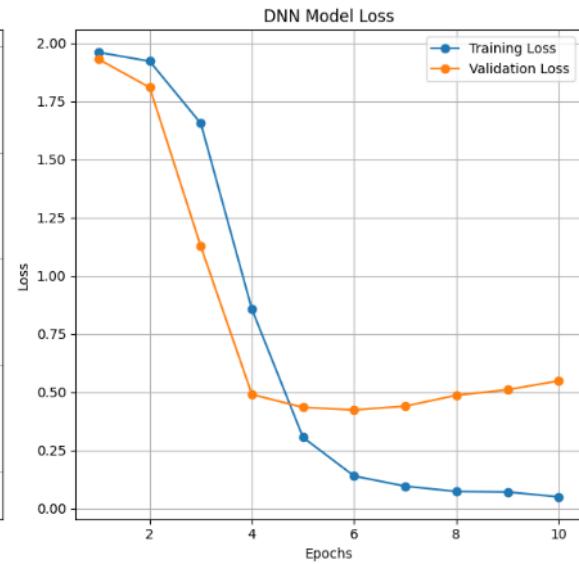
### 4.3.9 DNN Model



4.3.9.1 DNN Model Confusion Matrix



4.3.9.2 DNN Model Testing and Validation Accuracy



4.3.9.23 DNN Model Testing and Validation Loss

#### 4.4 Summary

We have demonstrated the models result, F1-score, Precision, Recall, Confusion matrix in this section. CNN model gave us the highest accuracy. Overall <sup>5</sup> deep learning model <sup>31</sup> is performing better than the machine learning models. We have shown training and validation accuracy for all <sup>1</sup> the deep learning <sup>79</sup> model applied here except Bangla Bert. We have also shown the training and validation loss of all <sup>1</sup> the deep learning <sup>79</sup> model applied here except Bangla Bert. Logistic regression, Random Forest Classifier, Multinomial Naive Bayes, SVM, Gradient Boosting Classifier, CNN, Bangla Bert, RNN, DNN accuracy respectively is 69%, 73%, 66%, 71%, 69%, 89%, 86%, 88%, 87%. The second highest accuracy we got from RNN.

# CHAPTER 5

## Engineering Standards and Design Challenges

### 5.1 Compliance with the standard

The Bangla text emotion detection system provides several benefits in terms of societal, environmental, and ethical impact while contributing to the need for sustainable practices. This section will discuss its broader implications and outline a sustainability plan.

#### 5.1.1 Communication Standards

For this thesis paper I had to consult with a psychologist for annotation validation.

Thus, I had to communicate with the psychologist. I had to discuss with my supervisor for conducting this thesis paper.

### 5.2 Impact on society, Environment & Sustainability

#### 5.2.1 Impact on Life

**Mental Health Support:** The system can help in the identification of emotions in Bangla text on social media or chat platforms, which could help identify signs of mental distress, anxiety, or depression; thus, enabling timely interventions that can save lives.

**Improved Communication:** It will allow machines to understand human emotions in

Bangla, hence allowing better interaction in applications such as chatbots and virtual assistants. This will help improve user satisfaction and experience.

**Educational Tools:** The system can aid in creating emotion-aware educational tools, enhancing emotional intelligence and language learning among Bangla-speaking individuals.

### **5.2.2 Impact on Society, Environment**

**Empowering Bangla Speakers:** This research addresses the linguistic gap in NLP technologies, providing Bangla-speaking communities with tools that resonate with their language and cultural context, fostering inclusivity in technology.

**Social Media Analysis:** This will, in turn, analyze the sentiments and emotions of the general public in Bangla texts and help organizations and policymakers to understand the societal mood in various issues and address those issues effectively.

### **5.2.3 Ethical Aspects**

**Data Privacy:** Emotion detection systems should handle user data in a secure and ethical manner. Data anonymization and data protection according to GDPR are crucial for building user trust.

**Bias Mitigation:** Making sure the model is fair to avoid amplifying biases present in the training dataset. This includes dealing with linguistic and cultural biases pertaining to Bangla.

**Transparency and Accountability:** Explanations of model predictions give transparency and allow users to trust the system, which is extremely important in applications like mental health or social media monitoring.

### **5.2.4 Sustainability Plan**

**Community Involvement:** Involve the Bangla-speaking communities in the improvement cycle of the system through contributions of annotated data and feedback. This will keep the system relevant and inclusive.

**Scalability and Accessibility:** Develop this system into a cloud-based service; this will give the benefit of scalability to users like institutions of education, health sectors, or social platforms.

### **5.3 Project Management and Financial Analysis**

For this project management we have calculated our budget. This is is the table given below:

#### **Primary Budget:**

TABLE 5.3.1 PRIMARY BUDGET

Category	Details	Cost (Taka)
Stake Holders	To verify the labeled data we need to visit a psychologist	1000 tk
Expert Help	Taking advices from an expert	2000 tk
Total		<b>3000 tk</b>

#### **Alternate Budget:**

TABLE 5.3.2 ALTERNATE BUDGET

Category	Details	Cost (Taka)
Stake Holders	To verify the data from my University psychologist	0
Expert help	Taking advices from an expert	1000
Total		<b>1000 tk</b>

## 5.4 Complex Engineering Problem

Complex engineering challenges often include related problems that need high-level technical knowledge, multidisciplinary collaboration, and creative solutions. The outcome of these kinds of problems is mostly unexpected, which needs rigorous research, design, and validation. This paper represents the first deep learning-based agriculture application for nut breed identification, integrating image processing with agricultural expertise to address real-world issues.

### 5.4.1 Complex Problem Solving

3  
TABLE 5.4.1.1 Mapping with complex problem solving.

EP1	EP2	EP3	EP4	EP5	EP6	EP7
Depth of Knowledge	Range of Conflicting Requirements	Depth of Analysis	Familiarity of Issues	Extent of Applicable Codes	Extent of Stakeholder Involvement	Interdependence
✓	✓	✓				✓

EP1 Depth of Knowledge: We had to go through multiple previous works to understand our work better and how to improve our work. Thus, we gained deep knowledge in this area.

Ep2 Range of Conflicting Requirements: I have studied in Computer Science and Engineering. I am implying machine learning and deep learning here. But I am doing research on emotion. That is psychology area.

Depth of Analysis: A vast amount of analysis had to be done to do this thesis paper.

Interdependence: As I have said earlier, I had to study previous papers to know my work in details.

### Mapping with Knowledge Profile for EP1

TABLE 5.4.1.2 Mapping with knowledge Profile (EP1)

<b>K3</b>	<b>K4</b>	<b>K5</b>	<b>K6</b>	<b>K8</b>
Engineering Fundamentals	Specialist Knowledge	Engineering Design	Engineering Practice	Research Literature
✓	✓	✓		✓

**K3 (Engineering Fundamentals):** Machine Learning and Deep learning models have been applied on the dataset.

**K4 (Specialist Knowledge):** An expert has verified the data.

**K5(Engineering Design):** In methodology the proposed diagram have been shown that is used to detect emotion from Bangla text.

**K8 (Research Literature):** To do this thesis paper we had to go through previous research papers.

### Mapping with Knowledge Profile for EP3

TABLE 5.4.1.3 Mapping with knowledge Profile (EP3)

<b>K2</b>	<b>K3</b>	<b>K4</b>	<b>K5</b>
Mathematics	Engineering Fundamentals	Specialist Knowledge	Engineering Design
✓	✓	✓	✓

**K2 (Mathematics):** To find out the accuracy we had to calculate the precision, recall, F1 score.

**K3 (Engineering Fundamentals):** <sup>46</sup> We have used deep learning and machine learning in this thesis paper.

**K4 (Specialist Knowledge):** We have taken an expert approval that our labeling data is correct.

**K5 (Engineering Design):** In methodology the proposed diagram have been shown that is used to detect emotion from Bangla text.

#### Mapping with Knowledge Profile for EP7

TABLE 5.4.1.4 Mapping with knowledge Profile (EP7)

K1	K2	K3	K4	K5	K6	K8
Natural Sciences	Mathematics	Engineering Fundamental s	Specialist Knowledge	Engineering Design	Engineering Practice	Research Literature
	✓	✓	✓	✓		✓

**K2 (Mathematics):** To find out the accuracy we had to calculate the precision, recall, F1 score.

**K3 (Engineering Fundamentals):** <sup>46</sup> We have used deep learning and machine learning in this thesis paper.

**K4 (Specialist Knowledge):** We have taken an expert approval that our labeling data is correct.

**K5 (Engineering Design):** In methodology the proposed diagram have been shown that is used to detect emotion from Bangla text.

**K8 (Research Literature):** To do this thesis paper we had to go through previous research papers.

#### 5.4.2 Engineering Activities

30

Table 5.4.2.1 Mapping with complex engineering activities

EA1	EA2	EA3	EA4	EA5
Range of resource s	Level of Interaction	Innovation	Consequences for society and environment	Familiarity
✓	✓		✓	✓

**EA1:** To conduct this paper variety of resources needed. Previous thesis paper, GPU, dataset.

**EA2:** To complete this thesis we had to come across supervisor and to verify the data we need to contact with a psychiatrist.

**EA4:** It has a contribution to society. To study about mental situation this study could help in the future.

#### 5.5 Summary

This thesis reflects such an interdisciplinary element of Emotion Recognition in Bangla Text by effectively combining the effort of machine learning with psychological knowledge to look into difficult engineering problems. In this thesis, careful examination of previously related works and their contribution, deep learning-based model, and methods, precision computation, recall, and F1-score, come into play to construct such a system design that a suggested engineering design diagram here is verified for its accuracy of labeling. The study utilized several resources such as GPUs, datasets, and expert collaboration. This work shows the societal benefit, especially in mental health research. The study also reflects interdisciplinary cooperation in resolving the competing demands from the field of psychology and computer science.

## Conclusion

### 6.1 Summary

In this thesis, we have explored the problem of emotion detection from Bangla text, with a focus on addressing the challenges associated with low-resource languages. Through a thorough review of existing methodologies and the implementation of state-of-the-art techniques, we have developed a robust framework for single-label emotion classification. By combining both classical machine learning algorithms and modern transformer-based models, we have showcased the efficacy of these approaches to capture the nuances of Bangla emotions. The results here indeed confirm that a high-quality dataset, advanced feature extraction techniques, and contextual embeddings improve the performance of Bangla emotion detection.

### 6.2 Limitation

Despite these promising results, there were some limitations to this study. Because most Bangla datasets originate from a few sites, including blogs or social media, the breadth and diversity of the datasets used were limited. This, in turn, limits the capacity of the models to be applied under different situations. Moreover, the model's capacity to predict all emotions equally was affected by the unequal distribution of emotion categories in the datasets. While transformer-based models did perform well, they came with some defects, especially in low-resource settings, since they are so computationally expensive and require much training data. Though we have collected data manually but it was not enough, we need more data to measure more accurately. We have detected a single emotion at a time. But one sentence can express multiple emotion at a time. Right now, we haven't developed an application.

### 6.1 Future Work

Future study can overcome these constraints by compiling bigger and more varied datasets covering a greater variety of Bangla text sources and emotional settings. The application of emotion detection algorithms may be further improved by developing domain-specific

datasets, such as those focusing on the customer service, medical, or educational fields. Further, investigating sophisticated architectures, such as pre-trained transformer models focused on Bangla, and using multimodal data, such as text combined with audio or visual inputs, could result in better performance and robustness. Another exciting future research direction is to expand into cross-lingual emotion recognition systems for Bangla and other South Asian languages.

## REFERENCE

- [2] D. G. Myers, *Theories of Emotion*, 7th ed. New York, NY, USA: Psychology Press, 2004.
- [60] S. A. Hosseini, "Classification of Brain Activity in Emotional States Using HOS Analysis," February 2012.
- [10] S. Azmin and K. Dhar, "Emotion Detection from Bangla Text Corpus Using Naïve Bayes Classifier," 2019 4th International Conference on Electrical Information and Communication Technology (EICT), Khulna, Bangladesh, 2019, pp. 1-6
- [9] R. R. Rushan, S. Hossain, S. S. Shovon, and M. A. Rahman, "Emotion Detection for Bangla Language.
- [14] H. A. Ruposh and M. M. Hoque, "A Computational Approach of Recognizing Emotion from Bengali Texts," 2019 5th International Conference on Advances in Electrical Engineering (ICAEE), Dhaka, Bangladesh, 2019, pp. 1-6.
- [3] A. K. Das, A. A. Asif, A. Paul, and M. N. Hossain, "Bangla Hate Speech Detection on Social Media Using Attention-Based Recurrent Neural Network," Journal of Intelligent Systems, vol. 30, no. 1, pp. 525–540, 2021. doi: 10.1515/jisys-2020-0060.
- [16] S. I. Khan, F. Bin Aziz, and M. U. Uddin, "Emotion Detection from Multilingual Text and Multi-Emotional Sentence Using Difference NLP Feature Extraction Technique and ML Classifier," International Journal of Advanced Networking and Applications, vol. 14, no. 3, pp. 5429–5435, 2022. ISSN: 0975-0290.
- [20] M. R. Faisal, A. M. Shifa, M. H. Rahman, M. A. Uddin, and R. M. Rahman, "Bengali & Banglali: A Monolingual Dataset for Emotion Detection in Linguistically Diverse Contexts," Data in Brief, Available online: <http://www.elsevier.com/locate/dib>, July 20, 2024.
- [22] M. A. Rahman and M. H. Seddiqi, "Comparison of Classical Machine Learning Approaches on Bangla Textual Emotion Analysis," arXiv preprint arXiv:1907.07826, Jul. 2019

- [10] T. Akter, M. S. Akter, T. Mahmud, D. Islam, M. S. Hossain, and K. Andersson, "Evaluating Machine Learning Methods for Bangla Text Emotion Analysis," 2024 Asia Pacific Conference on Innovation in Technology (APCIT), 2024, pp. 1–6, doi: 10.1109/APCIT62007.2024.10673544
- [11] S. A. Purba, S. Tasnim, M. Jabin, T. Hossen, and M. K. Hasan, "Document Level Emotion Detection from Bangla Text Using Machine Learning Techniques," 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), 2021, pp. 1–6
- [12] M. Mahmudun, M. T. Altaf, and S. Ismail, "Detecting Sentiment from Bangla Text using Machine Learning Technique and Feature Analysis," International Journal of Computer Applications, vol. 975, p. 8887
- [13] K. Sarkar, "Using Character N-gram Features and Multinomial Naive Bayes for Sentiment Polarity Detection in Bengali Tweets," in 2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT), 2018, pp. 1–4.
- [14] T. Rabeya, S. Ferdous, H. S. Ali, and N. R. Chakraborty, "A survey on emotion detection: A lexicon based backtracking approach for detecting emotion from Bengali text," in 2017 20th International Conference of Computer and Information Technology (ICCIT), 2017, pp. 1–7.
- [15] D. Das and S. Bandyopadhyay, "Labeling emotion in Bengali blog corpus--a fine grained tagging at sentence level," in Proceedings of the Eighth Workshop on Asian Language Resources, 2010, pp. 47–55.
- [16] H. Kavade, "A Logistic Regression Model to Predict Incident Severity Using the Human Factors Analysis and Classification System," M.S. thesis, Dept. Industrial Engineering, Clemson Univ., Clemson, SC, USA, Dec. 2009

# Comperative Analysis of Different Deep Learning Models to Classify Infectious Diseases from Chest Radiological Images

ORIGINALITY REPORT



PRIMARY SOURCES

- |   |  |     |
|---|--|-----|
| 1 | <b>dspace.daffodilvarsity.edu.bd:8080</b><br>Internet Source   | 4%  |
| 2 | O. Olabode, K. G. Akintola, and E. Adeyemi.<br>"Deep Learning Models for Classification of COVID-19 Severity Levels", Journal of Information Technology Innovations, 2024<br>Publication | 3%  |
| 3 | Submitted to Daffodil International University<br>Student Paper  | 2%  |
| 4 | Submitted to United International University<br>Student Paper  | 1 % |
| 5 | <b>www.mdpi.com</b><br>Internet Source   | 1 % |
| 6 | Abdelaziz Testas. "Distributed Machine Learning with PySpark", Springer Science and Business Media LLC, 2023<br>Publication  | 1 % |

7	Submitted to Liverpool John Moores University Student Paper	<1 %
8	Submitted to University of Hertfordshire Student Paper	<1 %
9	Avishek Das, Mohammed Moshiul Hoque, Omar Sharif, M. Ali Akber Dewan, Nazmul Siddique. "TEmoX: Classification of Textual Emotion using Ensemble of Transformers", IEEE Access, 2023 Publication	<1 %
10	thesai.org Internet Source	<1 %
11	Submitted to Universitas Sultan Ageng Tirtayasa Student Paper	<1 %
12	Tanjim Mahmud, Md. Alif Hossen Prince, Md. Hasan Ali, Mohammad Shahadat Hossain, Karl Andersson. "Enhancing Cybersecurity: Hybrid Deep Learning Approaches to Smishing Attack Detection", Systems, 2024 Publication	<1 %
13	Submitted to Higher Education Commission Pakistan Student Paper	<1 %
14	jurnal.atmaluhur.ac.id Internet Source	<1 %

15	Submitted to University of Surrey Student Paper	<1 %
16	<a href="http://www.ijana.in">www.ijana.in</a> Internet Source	<1 %
17	Suneeta Satpathy, Bijay Kumar Paikaray, Ming Yang, Arun Balakrishnan. "Sustainable Farming through Machine Learning - Enhancing Productivity and Efficiency", CRC Press, 2024 Publication	<1 %
18	Submitted to CSU Northridge Student Paper	<1 %
19	<a href="http://fastercapital.com">fastercapital.com</a> Internet Source	<1 %
20	Submitted to Bannari Amman Institute of Technology Student Paper	<1 %
21	Submitted to Universidade Nova De Lisboa Student Paper	<1 %
22	Abdullah Al Maruf, Fahima Khanam, Md. Mahmudul Haque, Zakaria Masud Jiyad, Firoj Mridha, Zeyar Aung. "Challenges and Opportunities of Text-based Emotion Detection: A Survey", IEEE Access, 2024 Publication	<1 %

Submitted to University of Teesside

23

&lt;1 %

24

[www.doubtly.in](http://www.doubtly.in)

&lt;1 %

25

Maliha Jahan, Istiak Ahamed, Md.  
Rayanuzzaman Bishwas, Swakkhar Shatabda.  
"Abusive Comments Detection in Bangla-  
English Code-mixed and Transliterated Text",  
2019 2nd International Conference on  
Innovation in Engineering and Technology  
(ICIET), 2019

&lt;1 %

Publication

26

"Advanced Network Technologies and  
Intelligent Computing", Springer Science and  
Business Media LLC, 2024

&lt;1 %

Publication

27

V. Sharmila, S. Kannadhasan, A. Rajiv Kannan,  
P. Sivakumar, V. Vennila. "Challenges in  
Information, Communication and Computing  
Technology", CRC Press, 2024

&lt;1 %

Publication

28

[digital.lib.washington.edu](http://digital.lib.washington.edu)

&lt;1 %

Internet Source

29

[wikimili.com](http://wikimili.com)

&lt;1 %

Internet Source

30

[www.coursehero.com](http://www.coursehero.com)

&lt;1 %

Internet Source

- 
- 31 "Proceedings of the 2nd International Conference on Big Data, IoT and Machine Learning", Springer Science and Business Media LLC, 2024 <1 %  
Publication
- 
- 32 Submitted to Fakultet elektrotehnike i računarstva / Faculty of Electrical Engineering and Computing <1 %  
Student Paper
- 
- 33 Submitted to Staffordshire University <1 %  
Student Paper
- 
- 34 medium.com <1 %  
Internet Source
- 
- 35 www.astesj.com <1 %  
Internet Source
- 
- 36 Submitted to King's College <1 %  
Student Paper
- 
- 37 Oladapo Oyebode, Ruth Ogubuike, Darlington Daniel, Rita Orji. "Chapter 32 Emotion Detection from Real-Life Situations Based on Journal Entries Using Machine Learning and Deep Learning Techniques", Springer Science and Business Media LLC, 2024 <1 %  
Publication
- 
- 38 docnum.univ-lorraine.fr

39

Moshiur Rahman Faisal, Ashrin Mobashira Shifa, Md Hasibur Rahman, Mohammed Arif Uddin, Rashedur M. Rahaman. "Bengali & Banglisch: A Monolingual Dataset for Emotion Detection in Linguistically Diverse Contexts", Data in Brief, 2024

Publication

40

Tamal Ahmed, Shawly Folia Mukta, Tamim Al Mahmud, Sakib Al Hasan, Md Gulzar Hussain. "Bangla Text Emotion Classification using LR, MNB and MLP with TF-IDF & CountVectorizer", 2022 26th International Computer Science and Engineering Conference (ICSEC), 2022

Publication

41

[link.springer.com](http://link.springer.com)

Internet Source

42

A. Pravin, C. Deepa. "Aggregate Plant Species Categorization with Deep CNN Based Positional Encoded Cross Attentional Mechanism", SN Computer Science, 2024

Publication

43

Submitted to London Metropolitan University

Student Paper

boto3.amazonaws.com

44	Internet Source	<1 %
45	jyx.jyu.fi Internet Source	<1 %
46	"Sentimental Analysis and Deep Learning", Springer Science and Business Media LLC, 2022 Publication	<1 %
47	Arafat, Mais Imad. "Detection of Wangiri Telecommunication Fraud Using Ensemble Learning", Princess Sumaya University for Technology (Jordan), 2022 Publication	<1 %
48	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1 %
49	Submitted to George Bush High School Student Paper	<1 %
50	nemertes.library.upatras.gr Internet Source	<1 %
51	trap.ncirl.ie Internet Source	<1 %
52	www.degruyter.com Internet Source	<1 %

- 53 Eilif Hjelseth, Sujesh F. Sujan, Raimar J. Scherer. "ECPPM 2022 – eWork and eBusiness in Architecture, Engineering and Construction", CRC Press, 2023 <1 %  
Publication
- 
- 54 Geeta Rana, R. K. Tailor, Vugar Abdullayev. "Data-Centric AI Solutions and Emerging Technologies in the Healthcare Ecosystem", CRC Press, 2023 <1 %  
Publication
- 
- 55 S.J. Xavier Savarimuthu, Sivakannan Subramani, Alex Noel Joseph Raj. "Artificial Intelligence for Multimedia Information Processing - Tools and Applications", CRC Press, 2024 <1 %  
Publication
- 
- 56 Submitted to The University of Memphis <1 %  
Student Paper
- 
- 57 [ijritcc.org](http://ijritcc.org) <1 %  
Internet Source
- 
- 58 [sadiaafripurba.github.io](https://sadiaafripurba.github.io) <1 %  
Internet Source
- 
- 59 [sinaahmadi.github.io](https://sinaahmadi.github.io) <1 %  
Internet Source
- 
- 60 [www.ir.juit.ac.in:8080](http://www.ir.juit.ac.in:8080) <1 %  
Internet Source

61	Submitted to Kingston University Student Paper	<1 %
62	dspace.bracu.ac.bd:8080 Internet Source	<1 %
63	export.arxiv.org Internet Source	<1 %
64	journalofbigdata.springeropen.com Internet Source	<1 %
65	section.iaesonline.com Internet Source	<1 %
66	www.frontiersin.org Internet Source	<1 %
67	www.packtpub.com Internet Source	<1 %
68	Alex Khang, Vugar Abdullayev, Babasaheb Jadhav, Shashi Kant Gupta, Gilbert Morris. "AI-Centric Modeling and Analytics - Concepts, Technologies, and Applications", CRC Press, 2023 Publication	<1 %
69	Koech, Gideon Kiprotich. "Investigation and Application of Artificial Intelligence Algorithms for Complexity Metrics Based Classification of Semantic Web Ontologies", Vaal University of Technology (South Africa), 2024 Publication	<1 %

70

Submitted to UNITEC Institute of Technology

Student Paper

<1 %

71

repository.up.ac.za

Internet Source

<1 %

72

Boße, Jonathan. "Forecasting Financial Asset Price Movements Using Convolutional Neural Networks – Application to the U.S. Financial Services Sector and Comparisson Across Industries", Universidade NOVA de Lisboa (Portugal), 2024

Publication

<1 %

73

Diptendu Sinha Roy, Mir Wajahat Hussain, K. Hemant Kumar Reddy, Deepak Gupta. "Healthcare-Driven Intelligent Computing Paradigms to Secure Futuristic Smart Cities", CRC Press, 2024

Publication

<1 %

74

Ramesh Chundi, Vishwanath R. Hulipalled, Jay Bharthish Simha. "NBLex: emotion prediction in Kannada-English code-switch text using naïve bayes lexicon approach", International Journal of Electrical and Computer Engineering (IJECE), 2023

Publication

<1 %

75

arxiv.org

Internet Source

<1 %

dokumen.pub

76	Internet Source	<1 %
77	repository.essex.ac.uk Internet Source	<1 %
78	semspub.epa.gov Internet Source	<1 %
79	www.irjmets.com Internet Source	<1 %
80	"Proceedings of International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications", Springer Science and Business Media LLC, 2021 Publication	<1 %
81	"Proceedings of the Future Technologies Conference (FTC) 2020, Volume 1", Springer Science and Business Media LLC, 2021 Publication	<1 %
82	Nasirul Mumenin, A.B.M. Kabir Hossain, Md. Arafat Hossain, Partha Pratim Debnath et al. "Screening depression among university students utilizing GHQ-12 and machine learning", Heliyon, 2024 Publication	<1 %
83	Srinivas Sethi, Bibhudatta Sahoo, Deepak Tosh, Suvendra Kumar Jayasingh, Sourav Kumar Bhoi. "Computing, Communication	<1 %

and Intelligence - A proceeding of ICCTCCI -  
2024", CRC Press, 2025

Publication

---

84	aclanthology.org	<1 %
85	conf.kln.ac.lk	<1 %
86	d197for5662m48.cloudfront.net	<1 %
87	downloads.hindawi.com	<1 %
88	ebin.pub	<1 %
89	html.pdfcookie.com	<1 %
90	tnsroindia.org.in	<1 %
91	uu.diva-portal.org	<1 %
92	www.iiis.org	<1 %
93	www.nits.ac.in	<1 %

- 94 Jay Liebowitz. "Data Analytics and AI", CRC Press, 2020 Publication <1 %
- 95 Punit Gupta, Dinesh Kumar Saini, Rohit Verma. "Healthcare Solutions Using Machine Learning and Informatics", CRC Press, 2022 Publication <1 %
- 96 Suman Lata Tripathi, Devendra Agarwal, Anita Pal, Yusuf Perwej. "Emerging Trends in IoT and Computing Technologies - Proceedings of Second International Conference on Emerging Trends in IoT and Computing Technologies – 2023 (ICEICT-2023), Goel Institute of Technology & Management Lucknow, India", CRC Press, 2024 Publication <1 %
- 97 Wu, Ruohan. "Artificial Intelligence (AI) Machine Learning Modelling of the Spatial Distribution of Groundwater Arsenic Hazard and its Attributable Health Risks", The University of Manchester (United Kingdom), 2024 Publication <1 %
- 98 Omar Faruqe, Mubassir Jahan, Md. Faisal, Md. Shahidul Islam, Riasat Khan. "Bangla Hate Speech Detection System Using Transformer-Based NLP and Deep Learning <1 %

# Techniques", 2023 3rd Asian Conference on Innovation in Technology (ASIANCON), 2023

Publication

---

99

Roy D Gregori Ayon, Md. Sanaullah Rabbi, Umme Habiba, Maoyejatun Hasana. "Bangla Speech Emotion Detection using Machine Learning Ensemble Methods", Advances in Science, Technology and Engineering Systems Journal, 2022

<1 %

Publication

---

Exclude quotes

Off

Exclude matches

Off

Exclude bibliography

Off