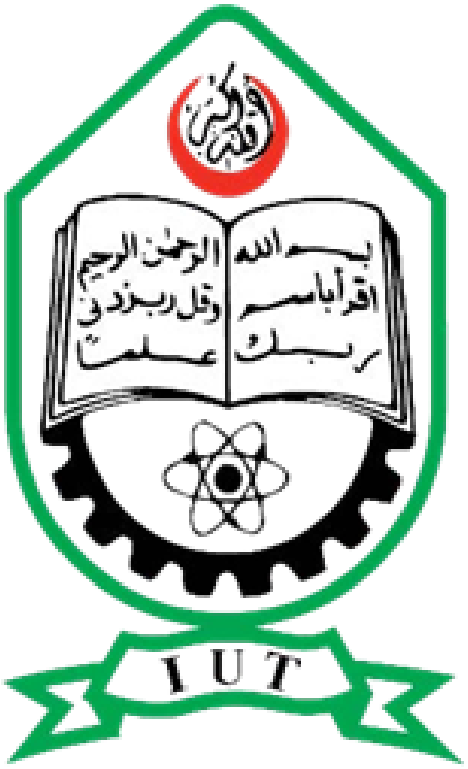


BACHELOR OF SCIENCE IN COMPUTER SCIENCE AND ENGINEERING

Detection of Social Media Cyberbullying in the Bengali Language : A Comparative Study

S.A.M Sajratul Yeaken Mollah Prangon
190042132
Anik Emtiaz Asif
190042133
Fayeaz-E-Abrar Taha
190042115

Department of Computer Science and Engineering
Islamic University of Technology
June, 2024

Contents

List of Figures

List of Tables

Declaration of Candidate

This is to certify that the work presented in this thesis is the outcome of the analysis and experiments carried out by **S.A.M Sajratul Yeaken Mollah Prangon**, **Anik Emtiaz Asif**, and **Fayeaz-E-Abrar Taha** under the supervision of **Faisal Hussain**, Assistant Professor, Department of Computer Science and Engineering, Islamic University of Technology, Dhaka, Bangladesh. It is also declared that neither this thesis nor any part of it has been submitted anywhere else for any degree or diploma. Information derived from the published and unpublished work of others have been acknowledged in the text and a list of references is given.

S.A.M Sajratul Yeaken Mollah Prangon
Student ID: 190042132
Date: June 04, 2024

Anik Emtiaz Asif
Student ID: 190042133
Date: June 04, 2024

Fayeaz-E-Abrar Taha
Student ID: 190042115
Date: June 04, 2024

Faisal Hussain
Assistant Professor
Department of Computer Science and Engineering
Islamic University of Technology (IUT)
Date: June 04, 2024

Acknowledgement

I want to express my sincere gratitude for appreciation for everyone who has supported and contributed to the successful finishing this thesis. In the first place, I owe my sincere appreciation to my supervisor, Faisal Hussain, whose expertise, constant encouragement, and constructive feedback have been invaluable throughout the research process. Their guidance has been a cornerstone in shaping the direction of my work and helping me navigate through the challenges with clarity and purpose.

I am also thankful to Islamic University of Technology for providing the essential resources, facilities, and a stimulating environment that enabled me to complete this thesis. I would also like to acknowledge the support of the faculty and staff, whose assistance has been immensely helpful at various stages of my academic journey.

My sincere thanks go to my friends and colleagues, who generously offered their time and valuable insights, helping me refine my ideas and broadening my perspective. Their feedback and encouragement were instrumental in pushing the boundaries of this research.

I would be remiss not to acknowledge my family, whose unwavering support, patience,Throughout this endeavor, my biggest sources of strength have been confidence and belief in myself. Their encouragement kept me motivated during the most challenging times.

Finally, to all those who contributed directly or indirectly to the success of this research, your support is deeply appreciated. This thesis would not have been possible without the collective efforts of all these individuals, and for that, I am sincerely grateful.

Abstract

As social media platforms have become more widely used, cyberbullying has become a major issue that cuts over linguistic and geographic barriers. Although significant efforts have been made to identify cyberbullying in major languages, little research has been done in this area regarding Bengali, despite its distinctive linguistic characteristics. An automated method for identifying instances of cyberbullying in Bengali social media content is presented in this thesis.

This study examines of current cyberbullying detection techniques and how well they work in the context of Bengali language usage.Our study includes feature extraction, classification, and data preprocessing, all of which are optimized to take into account the nuances of the Bengali language.

Important language elements that are typical of Bengali cyberbullying discourse are found and added to the feature extraction procedure. Additionally, semi-supervised learning approaches are used to improve classification performance by utilizing both labeled and unlabeled data in order to offset the lack of labeled datasets in Bengali.

Many experiments on real-world Bengali social media datasets are used to assess the efficacy of the suggested approach. performance metrics are compared amongst current approaches in terms of F1-score metrics, precision, and recall.

Chapter 1

Introduction

Social media is a important part of our lives, but it also badly effects our life by cyberbullying. Cyberbullying is bullying or harrassing someone in electronic media.Nowadays, cyberbullying is one of the growing concerns worldwide.While much research has focused on cyberbullying in languages like English, there is a significant gap in studies addressing this issue in Bengali language.
Bengali is spoken by over 230 million people of the world, mostly by Bangladeshi and West Bengal people. It is among the languages that are most often spoken worldwide. Despite the increasing use of social media among Bengali speakers, there are limited tools and methods for detecting cyberbullying in this language. The cultural and linguistic barrier may cause difficulties in understanding the slangs or bullies.
Our aim is to address these challenges by exploring effective strategies for detecting social media cyberbullying in Bengali. It covers the analysis of linguistic features, socio-cultural contexts, and the development of machine learning models tailored for Bengali. Our goal is to improve the detection and prevention of cyberbullying in the Bengali-speaking online community, promoting a safer and more inclusive digital environment.

1.1 Introduction to the Research Topic

Cyberbullying on social media is a growing concern that causes significant harm to victims, but research and detection tools in the Bengali language are limited. Bengali, which is spoken by over 230 million people, does not receive the same resources and attention as other major languages in this domain. The purpose of this study is to close this gap by developing effective methods for detecting cyberbullying in Bengali social media interactions. By focusing on linguistic nuances and utilizing advanced

machine learning techniques, we hope to develop tools that promote a safer online environment for Bengali speakers.

1.1.1 Motivations

Increasing Prevalence of Cyberbullying: Cyberbullying is a expanding problem on the platforms of social media, resulting in significant psychological harm. The increased use of social media by Bengali speakers necessitates addressing this issue in their native language.

Lack of Existing Research: Bengali, Even though it is the world's seventh most extensively spoken language, is underrepresented in cyberbullying detection research. Developing effective Bengali detection methods has the potential to fill a significant research gap.

Cultural and Linguistic Nuances: Understanding Bengali's distinct cultural and linguistic characteristics can help cyberbullying detection models perform more accurately. Addressing these nuances ensures that the solutions are culturally and contextually appropriate.

Developments in Technology: Natural language processing (NLP) and machine learning advances have created new opportunities for addressing language-specific challenges. Using these technologies can result in more sophisticated and effective detection systems for Bengali.

Social Impact: Encouraging the development of tools to identify and address cyberbullying can help create safer online spaces. One of society's most important objectives is to shield vulnerable groups, such as children and teenagers, from internet harassment.

1.1.2 Scope

Dataset Creation and Annotation: Building and compiling an extensive collection of social media exchanges in Bengali. Making certain that the dataset is diverse, accurately classified, and reflective of actual cyberbullying situations. **Linguistic Analysis:** Examining Bengali's morphological and syntactical structure to provide guidance for model creation. Addressing regional linguistic variances and dialectal variations within the dataset.

Model Development: Investigating several deep learning and machine learning models, with an emphasis on transformer-based architectures in particular. Experimenting with hyperparameter optimization, model selection, and fine-tuning specifically for Bengali text.

Evaluation Metrics: Defining and applying general evaluation measures (e.g., precision, recall, F1-score, confusion matrices) that go beyond accuracy. Addressing the dataset's class imbalance to guarantee impartial and equitable model performance.

Ethical and Privacy Considerations: Ensuring the model's impartiality and minimizing the propagation of bias from the training data. Observing moral principles and data protection laws, such the GDPR, when managing user-generated material.

Computational Efficiency: Ensuring the computational efficiency of the created models during both the training and inference stages. Concentrating on real-time detection skills to enable the models to be deployed feasibly.

Robustness and Generalization: Ensuring the models are resistant to hostile attacks and have good generalization to new data. Creating plans to deal with the deliberate avoidance techniques employed by cyberbullies.

Deployment and Impact Assessment: Preparing for the detection system's actual implementation on social media networks. Evaluating the system's effectiveness in lowering instances of cyberbullying among Bengali speakers.

1.1.3 Problem Statement

Cyberbullying is now a widespread problem on social networking sites that negatively impacts people's mental health and wellbeing, particularly that of younger users. Even with recent advances in natural language processing, it is still difficult to identify hostile and dangerous information in low-resource languages like Bengali. The purpose of this thesis is to close this gap by creating a reliable technique for identifying cyberbullying in Bengali writings. The study makes use of several transformer models that

have already been trained and assesses their effectiveness using a number of criteria, such as validation accuracy and F1-score. The goal is to determine which model and hyperparameter settings work best for this task, helping to create safer online spaces for people who speak Bengali.

1.1.4 Research Challenges

There are various obstacles in identifying cyberbullying on social media in Bengali. Given that Bengali is a low-resource language with little publicly available data and a history of bias and mislabeling, the quality and paucity of datasets present serious problems. Tokenization and modeling are made more difficult by the dialectal variances and morphological diversity of the language. Both significant computational resources and a great deal of experimentation are needed to choose and fine-tune the appropriate model. The use of extra metrics may be necessary since evaluation criteria such as accuracy and F1-score may not adequately describe model performance, particularly in the case of imbalanced datasets. Ensuring privacy compliance and preventing the propagation of bias are just two of the many ethical considerations that are crucial. Ultimately, practical deployment and reliability depend on the model's efficiency, generalization, and resilience to adversarial attacks.

1.1.5 Organization

This thesis is organized into five main chapters, each designed to address the essential aspects of the research on detecting cyberbullying in Bengali language social media.

- **Chapter 1 - Introduction:** This chapter introduces the research topic, discussing the significance of cyberbullying, particularly within the context of the Bengali language. It outlines the motivations, scope, problem statement, research challenges, and contributions of the study. This sets the foundation for the rest of the thesis, providing a clear understanding of the research objectives.
- **Chapter 2 - Related Works:** In this chapter, we review existing literature and studies that have addressed cyberbullying detection, particularly focusing on research related to other languages, machine learning models, and social media interactions. The chapter helps to situate our work within the broader context of cyberbullying research and highlights the gap this study aims to fill.
- **Chapter 3 - Stages of methodology:** This chapter details the methodological approach taken in the research. It explains the dataset selection, preprocessing techniques, and the machine learning models applied. The chapter also cov-

ers the training process, hyperparameter tuning, and evaluation metrics used to assess the model's performance in detecting cyberbullying in Bengali.

- **Chapter 4 - Results and Discussion:** This chapter presents the results obtained from the experiments conducted, including performance metrics such as accuracy, F1 score, and precision. It also includes a comparison of different models and their effectiveness in detecting cyberbullying in Bengali. The discussion section interprets these results in the context of the research objectives.
- **Chapter 5 - Conclusion:** The final chapter summarizes the research findings and discusses the implications of the results for future work. It highlights the contributions made by this thesis to the field of cyberbullying detection and suggests areas for further research to improve the detection systems for underrepresented languages like Bengali.

Chapter 2

Related Works

A number of researchers collectively highlight significant advancements in detecting and addressing online behaviors, particularly focusing on hope speech and cyberbullying across multiple languages and social media platforms. These studies underscore the importance of promoting equality, diversity, and inclusion through language technology, specifically by identifying and amplifying positive and supportive comments, referred to as hope speech. This is especially crucial for marginalized groups such as the LGBTQIA+ community, women in STEM, and persons with disabilities [7].

High precision and recall rates were attained in the hope speech detection study by using sophisticated machine learning and deep learning models, such as RoBERTa and XLM-RoBERTa [9]. This method stands out in particular because it makes use of a special multilingual dataset from social media sites that was meticulously annotated to guarantee ethical considerations by removing personal information. This endeavor not only generates a healthy online environment but also highlights the potential of language technologies in promoting social inclusion [7]. With models like GRU, BERT, and ELECTRA, notable progress has been achieved in the area of hate speech and offensive comment identification, especially in the Bengali language. When applied to Facebook datasets, these models demonstrated high accuracy rates; the BERT and ELECTRA models achieved approximately 85% [2], [10]. These developments are essential for upholding a civil and secure online community because they provide the efficient classification and mitigation of dangerous information.

[2], [10]. These developments are essential for upholding a civil and secure online community because they provide the efficient classification and mitigation of dangerous information.

Significant advancements in cyberbullying detection have also been made thanks to

transformer models including XLM-RoBERTa, Bengali DistilBERT, and Bangla BERT. XLM-RoBERTa achieved the highest accuracy and F1-scores when these models were applied to huge datasets of Facebook comments in Bangla [10], [13]. Strong cyberbullying detection systems must be created in order to shield people from online abuse and to safeguard their mental and emotional health.

Cyberbullying's psychological effects are well-established; studies have shown that it can cause severe emotional anguish, anxiety, sadness, and low self-esteem [3]–[5]. These findings highlight the critical necessity for victim support networks and efficient treatments. School nurses can help victims, students, and families who are impacted by cyberbullying by implementing practical measures that give them the knowledge and resources they need to fight this problem [6].

Additionally, in order to inform interventions and policies, insights into how students perceive cyberbullying on social media platforms have been investigated [1], [17]. From the viewpoint of the students, this study offers a deeper comprehension of the frequency, characteristics, and effects of cyberbullying. Undergraduate students' protective methods to protect themselves from cyberbullying have also been studied, providing important information about how effective these tactics are [11].

The fundamental dynamics and patterns of cyberbullying on social media platforms have been revealed through the application of sophisticated data analysis tools. In order to create focused interventions, it is essential to investigate the ways in which cyberbullying appears and proliferates online [8]. Furthermore, research has examined the particular difficulties faced by victims of cyberbullying on mobile devices, emphasizing the necessity of tailored strategies to combat mobile cyberbullying [18].

All things considered, these studies offer a thorough summary of contemporary approaches and interventions meant to promote online positivity and stop cyberbullying. Together, they support marginalized communities, make the internet a safer and more welcoming place, and enhance the general wellbeing of those impacted by harmful online conduct.

Chapter 3

Stages of Our Study

- **Dataset Selection:**
  We have used the "BAD-Bengali-Aggressive-Text-Dataset," which contains a collection of Bengali social media texts annotated for aggressive and non-aggressive content. This dataset was sourced from a public GitHub repository. It provides a balanced distribution of aggressive and non-aggressive text, making it suitable for training machine learning models for the task of detecting aggression in Bengali social media interactions.
  The dataset was pre-processed to remove noise such as URLs, special characters, and extra spaces, while maintaining the core text features important for identifying aggressive speech.
- **Model Selection:**
  We selected and experimented with five different machine learning models tailored to natural language processing (NLP), specifically focused on Bengali language text:
  - `sagorsarker/bangla-bert-base`: A BERT-based model pre-trained on a large corpus of Bengali text, designed to handle Bengali language nuances.[15]
  - `drygegedsgryeuj/test`: A test model used for benchmarking. This model explores lightweight architectures for low-resource scenarios.[14]
  - `neuropark/sahajBERT`: Another BERT-based model with optimizations specific to Bengali text processing.[19]
  - `bangla-speech-processing/BanglaASR`: Primarily designed for speech recognition, we adapted this model for text-based classification tasks.[12]

- `bangla-sentence-transformer`: A transformer model optimized for producing sentence embeddings in Bengali, useful for various NLP tasks, including classification.[16]

Each model was selected based on its ability to handle complex sentence structures and semantic nuances of the Bengali language.

- **Hyperparameter Tuning:**
  We experimented with two key hyperparameters: the number of epochs and learning rates. Tuning these hyperparameters allowed us to optimize the models' performance:
  - **Epochs:** We trained each model for 3 and 5 epochs. Increasing the number of epochs allows the model to learn more from the data but can lead to overfitting if too high.
  - **Learning Rates:** Two learning rates were tested: 5e-5 and 1e-5. The learning rate determines how much the model should be altered in reaction to the

    estimated inaccuracy every time the weights in the model are changed. A lower learning rate can result in better performance, but it may require more epochs to converge.

  By fine-tuning these hyperparameters, we aimed to find the optimal balance between model performance and training time.

- **Training and Evaluation:**
  Each model was trained on the preprocessed dataset, and we followed these steps for each combination of model and hyperparameter settings:
  - **Data Loading:** The dataset was split into training and validation sets. DataLoaders were created to handle batch processing and shuffling to ensure robust training.
  - **Tokenization:** The text data was tokenized using a pre-trained tokenizer specific to each model. Tokenization breaks the text into words, subwords, or symbols that the model can process.
  - **Training:** Each model was trained using the selected hyperparameters. For every epoch, the model was updated based on the loss function, which measures the error in the model's predictions.
  - **Evaluation:** After training, we evaluated each model on the validation

set using two key metrics:

* **Validation Accuracy:** The percentage of correct predictions made by the model on the validation set.
* **F1 Score (Macro Averaged):** A measure that combines both precision and recall, giving a single score to evaluate the model's performance. The macro-averaged F1 score treats each class equally, regardless of the class distribution.

The evaluation results were used to compare model performance and select the most effective model for detecting aggression in Bengali text.

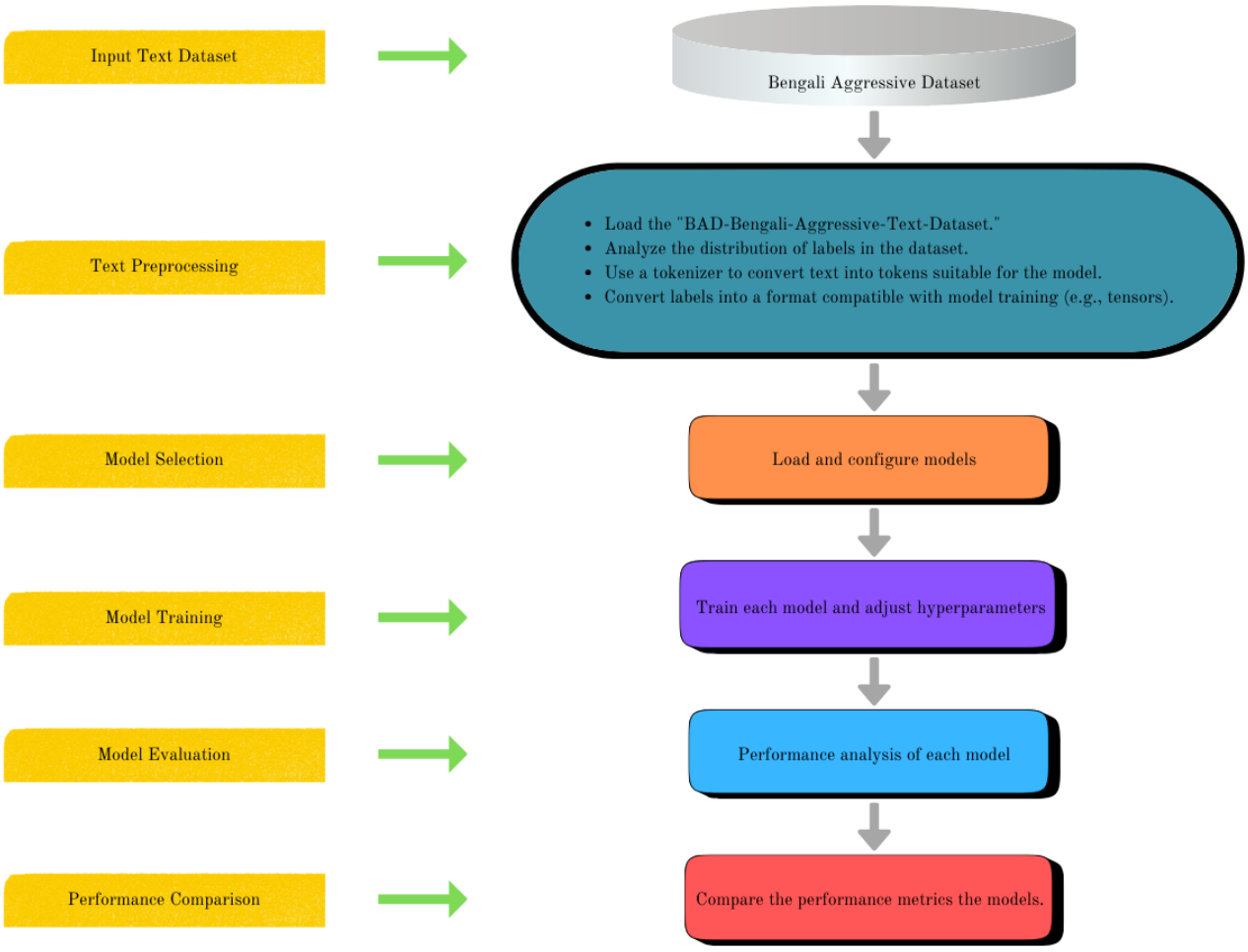


Figure 3.1: Methodology

The Figure has provided outlines the methodology for detecting aggressive or abusive comments in Bengali using various transformer models.

- **Input Text Dataset**
  Description: The process begins with the input of the text dataset, specifically the "Bengali Aggressive Dataset."
  Action: Load the dataset which contains Bengali text comments labeled for aggressiveness.

- **Text Preprocessing**
  Description: Preprocessing the text data is crucial for preparing it for model training.
  Steps: Analyze the Distribution of Labels: Examine how the labels (aggressive vs. non-aggressive) are distributed within the dataset.
  Tokenization: Convert the raw text into tokens, which are the basic units that the model will process. This involves splitting the text into words or subwords.
  Label Conversion: Transform the labels into a format that is compatible with model training, such as tensors in machine learning frameworks.

- **Model Selection**
  Description: Choose appropriate models that are suitable for the task of detecting aggressive speech.
  Action: Load and configure various models for comparison. This might include popular transformer models like BERT, RoBERTa, or specialized versions for the Bengali language.

- **Model Training**
  Description: Train each selected model on the preprocessed dataset.
  Steps: Train Each Model: Use the dataset to train the models, adjusting the hyperparameters (e.g., learning rate, batch size) to optimize performance.
  Adjust Hyperparameters: Fine-tune the hyperparameters to improve the models' ability to accurately classify the text as aggressive or non-aggressive.

- **Model Evaluation**
  Description: Evaluate the performance of each trained model.
  Steps: Performance Analysis: Assess each model based on key metrics such as accuracy, precision, recall, and F1 score. This step involves validating the models on a separate validation dataset to ensure they generalize well to new, unseen data.

- **Performance Comparison**
  Description: Compare the performance metrics of all the trained models to determine the best one.
  Steps: Compare Metrics: Analyze the results of the performance evaluation and compare the models based on the collected metrics. This comparison helps in identifying which model performs best in detecting aggressive comments in Bengali.

Each step in this methodology is designed to ensure a comprehensive and systematic approach to training, evaluating, and selecting the best-performing model for the task of detecting aggressive speech in Bengali social media comments. The process emphasizes the importance of thorough preprocessing, careful model selection, rigorous training, and detailed evaluation to achieve optimal results.

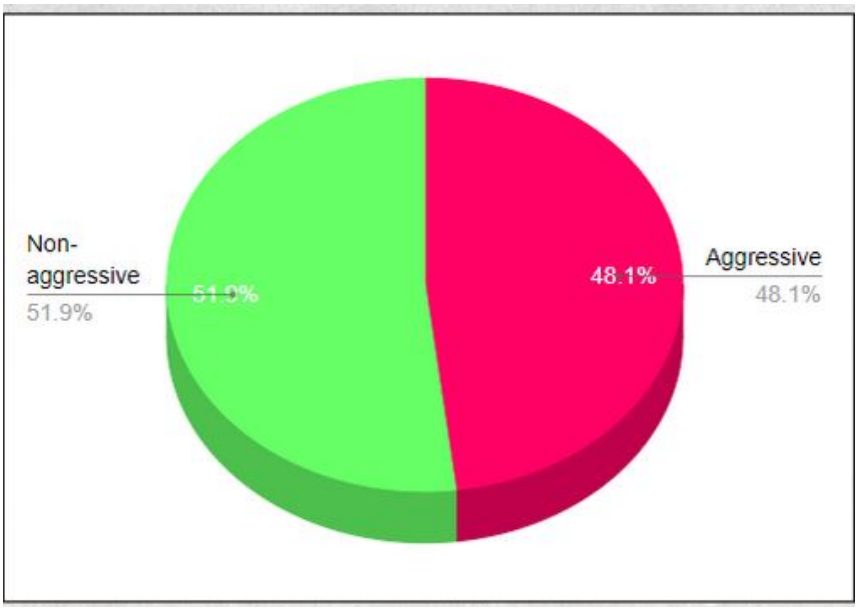


Figure 3.2: Percentage of Aggressive and Non-aggressive Texts

Non-aggressive: Proportion: 51.9 percentage.This class forms the majority of the dataset.Aggressive: 48.1 percentage. This class forms the minority of the dataset, but it's nearly balanced with the Non-aggressive class. Class Distribution: The distribution between the two classes is quite balanced, with "Non-aggressive" having a slight majority. This balance is beneficial for model training because it reduces the risk of the model becoming biased towards one class.Implications for Model Training:(Balanced Classes)-Since the classes are almost evenly distributed, the model will have a fair representation of both classes during training, leading to better generalization.Performance Metrics: With such a balanced dataset, performance metrics like accuracy, precision, recall, and F1 score will provide a more accurate reflection of the model's performance.Further

Evaluation: To gain deeper insights into the model's performance, we should look at confusion matrices, precision-recall curves, and ROC curves. These will help us understand how well the model distinguishes between the "Aggressive" and "Non-aggressive" classes.

```
model.train()
    for epoch in range(NUM_EPOCHS):
        for batch in train_dataloader:
            input_ids = batch['input_ids'].to(device)
            attention_mask = batch['attention_mask'].to(device)
            labels = batch['label'].to(device)
            optimizer.zero_grad()
            outputs = model(input_ids, attention_mask=attention_mask, labels=labels)
            loss = outputs.loss
            loss.backward()
            optimizer.step()
        print(f"Epoch {epoch+1} completed with loss: {loss.item()}")
```

Figure 3.3: Pseudo Code of Model Training

Pseudo Code Explanation

The provided code represents a typical training loop for machine learning models. It starts by setting the model to training mode, ensuring that all layers are active for learning. The outer loop iterates over multiple epochs, each representing a complete pass through the training dataset. Inside this loop, the data is processed in batches, which helps manage memory efficiently and accelerates computation. For each batch, the input data is transferred to the appropriate device (e.g., CPU or GPU), and the gradients from the previous iteration are reset to prevent accumulation. The model performs a forward pass to generate predictions, which are then compared to the ground truth labels to compute the loss using a loss function. A backward pass follows, where gradients are calculated and propagated through the model. Finally, the optimizer updates the model parameters based on these gradients to minimize the loss. After each epoch, the code outputs the loss value, allowing the user to monitor the training progress.

The code snippet represents a training loop that:

Sets the model to training mode, iterates over multiple epochs and batches, moves data to the appropriate device, performs forward and backward passes to compute and propagate loss, updates model parameters based on the computed gradients, prints the loss value after each epoch to monitor training progress.

Chapter 4

Results and Discussion

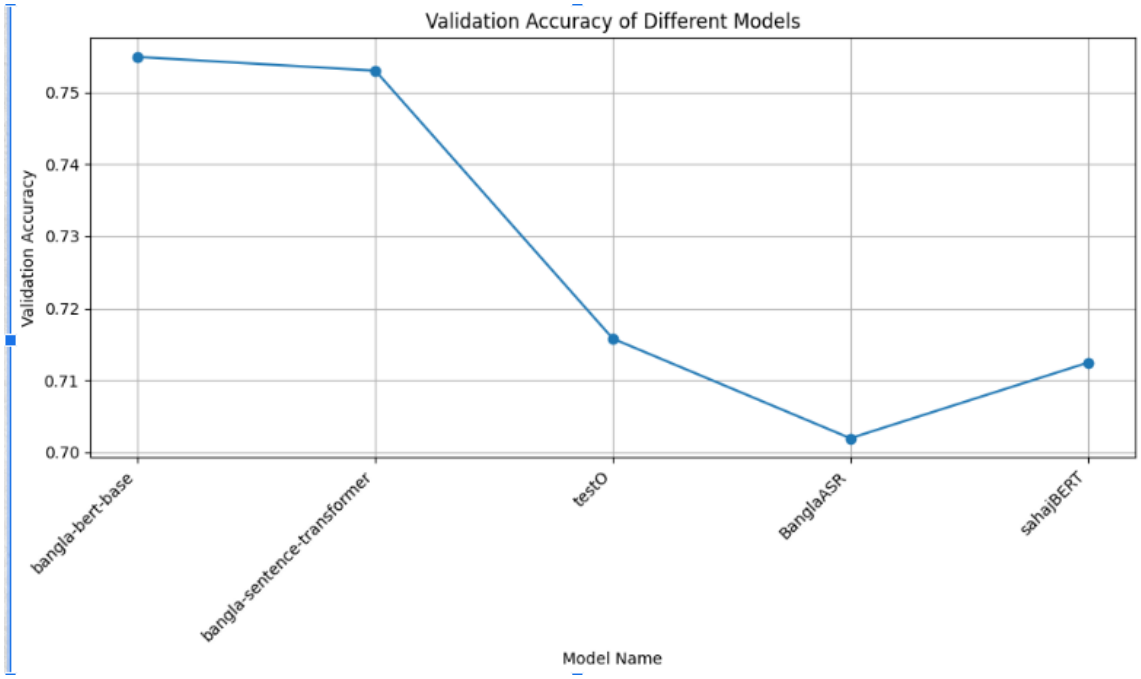


Figure 4.1: Accuracy of Different Models

The graph you have provided is a line plot that represents the validation accuracy of different models. Here's a detailed description:

Axes :

X-Axis (Model Name): This axis lists the names of the different models used for validation.

The models listed are:

Model Descriptions

bangla-bert-base

bangla-bert-base is a pre-trained BERT-based model designed specifically for processing Bengali text. It leverages a large corpus of Bengali language data to capture linguistic nuances, including morphology and syntax, making it well-suited for tasks such as text classification and sentiment analysis.

bangla-sentence-transformer

bangla-sentence-transformer is a transformer-based model optimized for generating sentence embeddings in Bengali. These embeddings capture semantic relationships between sentences, making the model effective for downstream tasks like text similarity and classification.

textO

textO is a lightweight and experimental model tailored for benchmarking in low-resource scenarios. While its primary purpose is to explore efficient architectures, it also provides baseline performance for Bengali language tasks.

BanglaASR

BanglaASR is primarily an Automatic Speech Recognition (ASR) model for Bengali. Although designed for speech-to-text tasks, it has been adapted for text classification by leveraging its ability to process the Bengali language effectively.

sahajBERT

sahajBERT is another BERT-based model with specific optimizations for Bengali text processing. Its design emphasizes simplicity and efficiency, making it suitable for applications requiring fast and accurate language understanding.

Y-Axis (Validation Accuracy): This axis represents the validation accuracy of each model, ranging from 0.70 to 0.76.

Data Points:

bangla-bert-base: This model has the highest validation accuracy, slightly above 0.75. bangla-sentence-transformer: This model has a validation accuracy close to that of bangla-bert-base, but very slightly lower. textO: This model has a lower validation accuracy, around 0.72. BanglaASR: This model has the lowest validation accuracy, around 0.71. sahajBERT: This model has a validation accuracy higher than BanglaASR but lower than textO, approximately 0.71.

Trend:

The line plot shows a decreasing trend in validation accuracy from bangla-bert-base to BanglaASR, with a slight increase for sahajBERT.

Interpretation:

Highest Accuracy: bangla-bert-base and bangla-sentence-transformer models are the most accurate based on this validation accuracy metric.
Lowest Accuracy: BanglaASR has the lowest validation accuracy among the models tested.
General Trend: There is a noticeable drop in validation accuracy from the first two models to the subsequent ones, with a minor recovery at the end with sahajBERT.

This plot is useful for comparing the performance of different models in terms of validation accuracy and making decisions on which model might be the best for a specific task.

The provided graph is a line plot illustrating the validation accuracy of different models. Here's a concise description:

Axes:

X-Axis (Model Name): Lists the names of five models:

Y-Axis (Validation Accuracy): Ranges from 0.69 to 0.74.

Data Points:

bangla-bert-base: Has a validation accuracy of about 0.70.
bangla-sentence-transformer: Shows a higher validation accuracy of approximately 0.73.

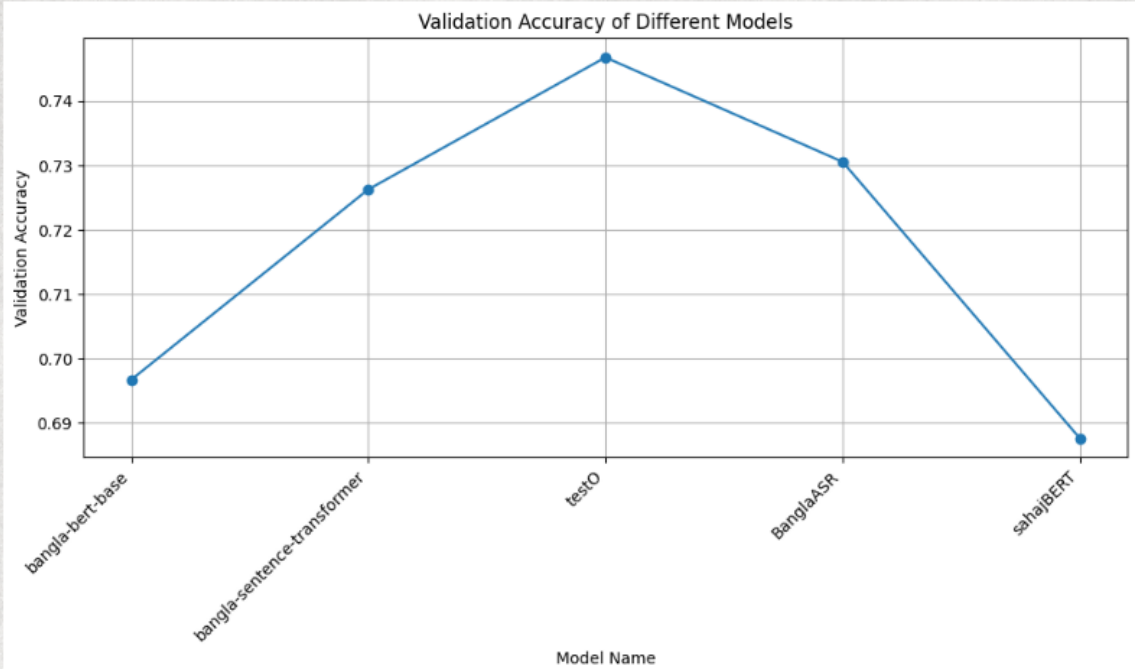


Figure 4.2: Accuracy of Different Models

textO: Peaks with the highest validation accuracy around 0.74.
BanglaASR: Drops to a validation accuracy near 0.72.
sahajBERT: Has the lowest validation accuracy, around 0.69.

Trend:

The plot shows an increase in validation accuracy from bangla-bert-base to textO, which has the highest accuracy.
After textO, the validation accuracy decreases, with sahajBERT having the lowest accuracy among the models.

Summary:

textO is the most accurate model.
sahajBERT is the least accurate.
There is an upward trend in accuracy until textO, followed by a downward trend.

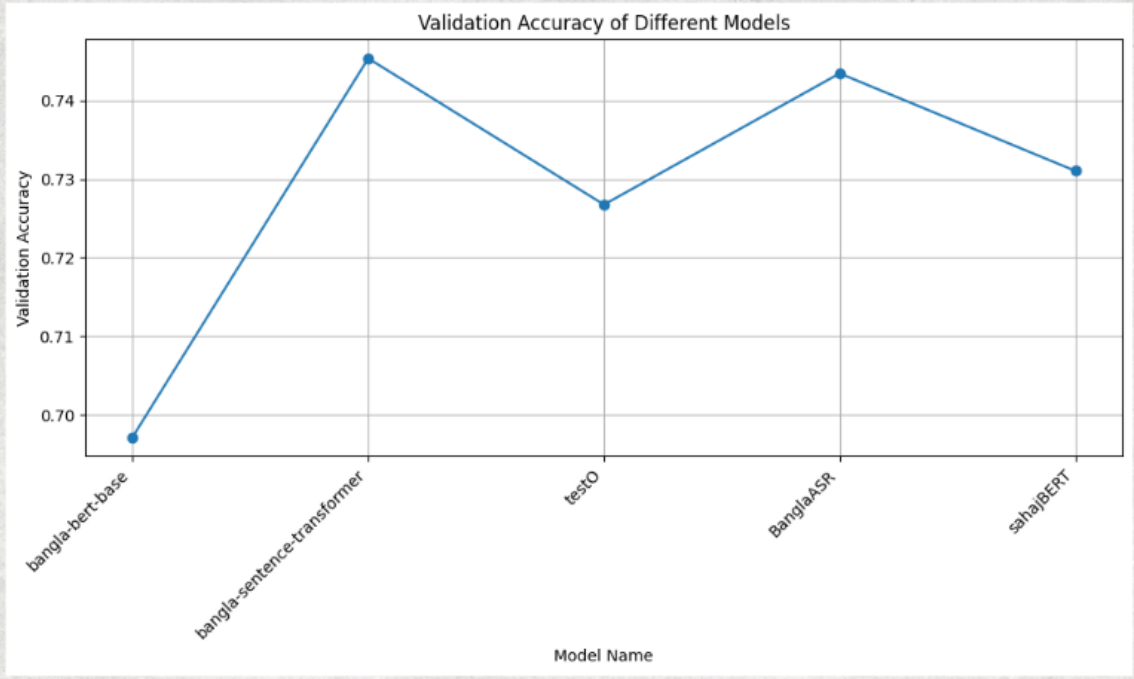


Figure 4.3: Accuracy of Different Models

Axes :

X-Axis (Model Name): Lists the names of five models: Y-Axis (Validation Accuracy): Ranges from 0.70 to 0.74.

Data Points:

bangla-bert-base: Starts with a validation accuracy of about 0.70.
bangla-sentence-transformer: Peaks with the highest validation accuracy, slightly above 0.74.
textO: Drops to around 0.73.
BanglaASR: Increases again, reaching slightly above 0.74.
sahajBERT: Decreases to around 0.72.

Trend:

The plot shows a rise in validation accuracy from bangla-bert-base to bangla-sentence-transformer. It then decreases at textO, rises again at BanglaASR, and finally drops at sahajBERT.

Summary:

Highest Accuracy: bangla-sentence-transformer and BanglaASR have the highest validation accuracies, slightly above 0.74.
Lowest Accuracy: bangla-bert-base has the lowest validation accuracy, around 0.70.
There is a fluctuating trend in validation accuracy among the models, with peaks at bangla-sentence-transformer and BanglaASR.

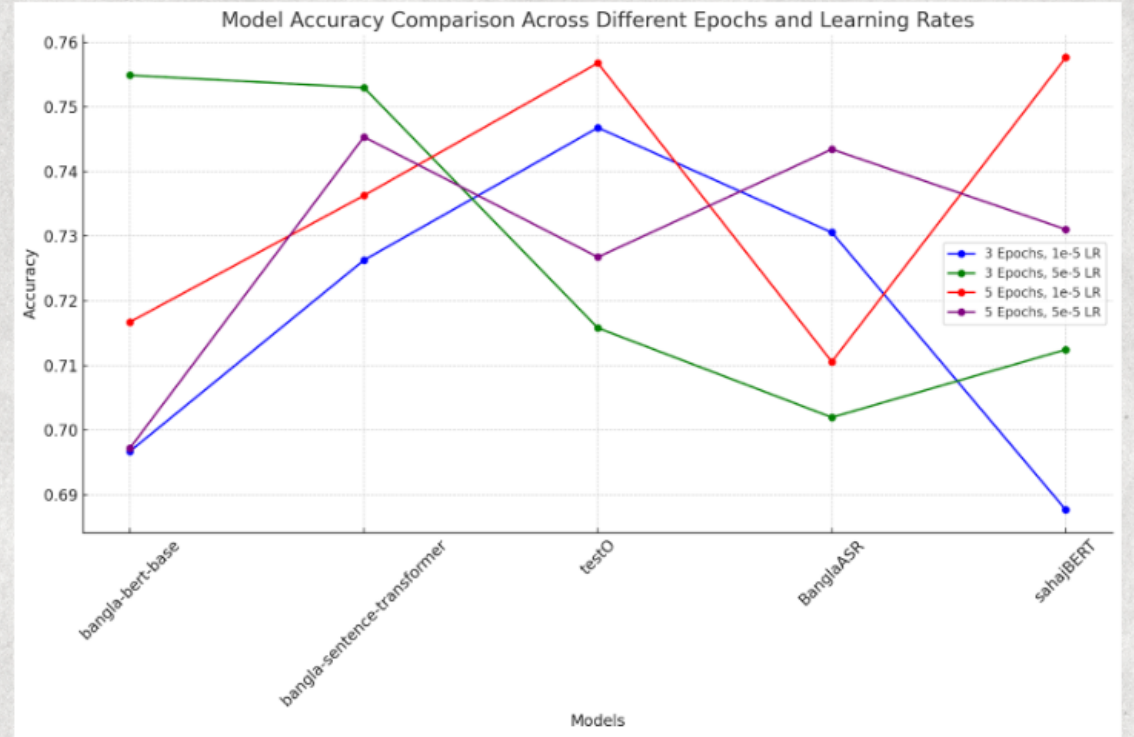


Figure 4.4: Comparison of Model Accuracies

The graph titled "Model Accuracy Comparison Across Different Epochs and Learning Rates" illustrates the accuracy of various models trained under different conditions.
The x-axis represents all the models
The y-axis represents the accuracy, ranging from 0.69 to 0.76. The graph includes four lines, each representing a different combination of epochs and learning rates:

Blue line: 3 Epochs, 1e-5 Learning Rate (LR)
Green line: 3 Epochs, 5e-5 LR
Red line: 5 Epochs, 1e-5 LR
Purple line: 5 Epochs, 5e-5 LR

Key observations:

The accuracy varies across different models and training conditions. "sahajBERT" with 5 Epochs, 1e-5 LR (red line) has the highest accuracy. "bangla-bert-base" consistently performs well across different conditions, especially with 5 Epochs, 1e-5 LR. "testO" and "BanglaASR" exhibit more variation in accuracy based on the training conditions. This comparison helps in understanding which combinations of epochs and learning rates yield the best performance for each model.

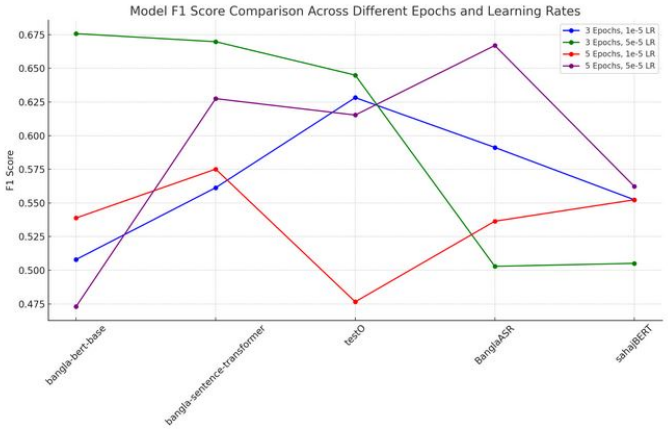
Table 4.1: Accuracy Comparison

Accuracy Comparison:

Model	3 Epochs, 1e-5 LR	3 Epochs, 5e-5 LR	5 Epochs, 1e-5 LR	5 Epochs, 5e-5 LR
bangla-bert-base	**0.6967**	**0.7549**	0.7153	0.6972
bangla-sentence-transformer	0.7263	**0.7530**	**0.7039**	0.7454
testO	**0.7468**	0.7158	**0.6853**	0.7268
BanglaASR	0.7306	0.7020	**0.6962**	**0.7434**
sahajBERT	**0.6876**	0.7124	**0.6876**	**0.7310**

Table 4.2: F1 Score Comparison

F1 Score Comparison:



Let's break down the key elements:

- Title and Labels:

- Title: The chart is titled "Model F1 Score Comparison Across Different Epochs and Learning Rates," indicating that it compares the performance of different models under various training conditions.
- Y-Axis: Represents the F1 score, a metric that considers both precision and recall, providing a single score to evaluate the model's performance.
- X-Axis: Lists different models: bangla-bert-base bangla-sentence-transformer textO BanglaBERT sahajBERT
- Legends:

  Colors: Each line represents a different training configuration: Blue Line: 3 Epochs, 1e-5 Learning Rate (LR) Green Line: 3 Epochs, 5e-5 LR Red Line: 5 Epochs, 1e-5 LR Purple Line: 5 Epochs, 5e-5 LR
- Analysis:
- bangla-bert-base: Highest F1 score achie0000ved with 3 Epochs, 1e-5 LR (Blue line). Lower F1 scores with other configurations, indicating that the model performs best with a lower learning rate and fewer epochs.
- bangla-sentence-transformer: Best performance with 5 Epochs, 5e-5 LR (Purple line). The model shows significant improvement with higher learning rates and more epochs, suggesting it benefits from more extensive training.
- textO: Highest F1 score achieved with 5 Epochs, 5e-5 LR (Purple line). The scores vary significantly, indicating sensitivity to training configurations.
- BanglaBERT: Optimal performance with 3 Epochs, 1e-5 LR (Blue line). The model shows a steep decline with other configurations, especially with 5 Epochs and 1e-5 LR (Red line).
- sahajBERT: Performs best with 3 Epochs, 1e-5 LR (Blue line). Other configurations yield lower scores, showing less sensitivity to changes compared to other models.
- Key Observations:
- Consistency in Performance: bangla-bert-base, BanglaBERT, and sahajBERT tend to perform best with 3 Epochs and a 1e-5 learning rate. Sensitivity to Training Configurations: Models like bangla-sentence-transformer and textO show more variation across different epochs and learning rates, suggesting they may require more tuning to achieve optimal performance.

Higher F1 Scores with Certain Configurations: For most models, the best configurations appear to be 3 Epochs with a lower learning rate (1e-5), but bangla-sentence-transformer and textO benefit from longer training (5 Epochs) and a higher learning rate (5e-5).

Table 4.3: F1 Score Comparison

F1 Score Comparison:

Model	3 Epochs, 1e-5 LR	3 Epochs, 5e-5 LR	5 Epochs, 1e-5 LR	5 Epochs, 5e-5 LR
bangla-bert-base	0.5079	**0.6758**	0.5388	**0.4730**
bangla-sentence-transformer	**0.5613**	**0.6698**	0.5751	0.6275
testO	0.6283	**0.6450**	**0.4765**	0.6153
BanglaASR	0.5912	**0.5029**	0.5364	**0.6670**
sahajBERT	0.5523	**0.5051**	0.5523	**0.5622**

Chapter 5

Conclusion

The Conclusions chapter of a thesis book is a critical section that summarizes our research, highlights the significance of our findings, and suggests potential future research directions. Here are the key elements to include in the Conclusions chapter:

- The bangla-sentence-transformer consistently delivers strong performance, with its lowest accuracy being 72.628
- The testO model, despite being one of the worse-performing models overall, achieves its best performance with a lower number of epochs and learning rate.
- bangla-bert-base and bangla-sentence-transformer achieve their highest accuracies (75.489
- Both sahajBERT and bangla-bert-base fall short of the industry standard with accuracies below 70

These findings highlight the importance of selecting appropriate training settings to optimize model performance, showing that the bangla-sentence-transformer is the most robust model overall, while other models may require specific conditions to perform at their best.

References

[1] A. Akrim, "Student perception of cyberbullying in social media," *Aksaqila Jabfung*, 2022.

[2] T. T. Aurpa, R. Sadik, and M. S. Ahmed, "Abusive bangla comments detection on facebook using transformer-based deep learning models," *Social Network Analysis and Mining*, vol. 12, no. 1, p. 24, 2022.

[3] S. Batool, R. Yousaf, and F. Batool, "Bullying in social media: An effect study of cyber bullying on the youth," *Pakistan Journal of Criminology*, vol. 9, no. 4, p. 119, 2017.

[4] A. Bozyiğit, S. Utku, and E. Nasibov, "Cyberbullying detection: Utilizing social media features," *Expert Systems with Applications*, vol. 179, p. 115 001, 2021.

[5] E. Byrne, J. A. Vessey, and L. Pfeifer, "Cyberbullying and social media: Information and interventions for school nurses working with victims, students, and families," *The Journal of School Nursing*, vol. 34, no. 1, pp. 38–50, 2018.

[6] M. A. Carter, "Protecting oneself from cyber bullying on social media sites–a study of undergraduate students," *Procedia-Social and Behavioral Sciences*, vol. 93, pp. 1229–1235, 2013.

[7] B. R. Chakravarthi, "Multilingual hope speech detection in english and dravidian languages," *International Journal of Data Science and Analytics*, vol. 14, no. 4, pp. 389–406, 2022.

[8] D. Chatzakou, I. Leontiadis, J. Blackburn, *et al.*, "Detecting cyberbullying and cyberaggression in social media," *ACM Transactions on the Web (TWEB)*, vol. 13, no. 3, pp. 1–51, 2019.

[9] R. R. Dalvi, S. B. Chavan, and A. Halbe, "Detecting a twitter cyberbullying using machine learning," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, IEEE, 2020, pp. 297–301.

[10] M. I. H. Emon, K. N. Iqbal, M. H. K. Mehedi, M. J. A. Mahbub, and A. A. Rasel, "Detection of bangla hate comments and cyberbullying in social media using

nlp and transformer models," in *International Conference on Advances in Computing and Data Sciences*, Springer, 2022, pp. 86–96.

[11] A. Görzig and L. A. Frumkin, "Cyberbullying experiences on-the-go: When social media can become distressing," *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, vol. 7, no. 1, 2013.

[12] F. Hassan, M. R. A. Kotwal, and M. N. Huda, "Bangla asr design by suppressing gender factor with gender-independent and gender-based hmm classifiers," in *2011 World Congress on Information and Communication Technologies*, 2011, pp. 1276–1281. DOI: 10.1109/WICT.2011.6141432.

[13] A. Ishmam and S. Sharmin, "Hateful speech detection in public facebook pages for the bengali language," Dec. 2019, pp. 555–560. DOI: 10.1109/ICMLA.2019.00104.

[14] H. Kelejian, "A spatial j-test for model specification against a single or a set of non-nested alternatives," *AStA Wirtschafts- und Sozialstatistisches Archiv*, vol. 1, pp. 3–11, Feb. 2008. DOI: 10.1007/s12076-008-0001-9.

[15] M. Kowsher, A. A. Sami, N. J. Prottasha, M. S. Arefin, P. K. Dhar, and T. Koshiba, "Bangla-bert: Transformer-based efficient model for transfer learning and language understanding," *IEEE Access*, vol. 10, pp. 91 855–91 870, 2022. DOI: 10.1109/ACCESS.2022.3197662.

[16] H. Shahgir and K. Sayeed, *Bangla grammatical error detection using t5 transformer model*, Preprint available on arXiv, Mar. 2023. DOI: 10.48550/arXiv.2303.10612. arXiv: 2303.10612 [cs.CL].

[17] S. Singh, V. Thapar, and S. Bagga, "Exploring the hidden patterns of cyberbullying on social media," *Procedia Computer Science*, vol. 167, pp. 1636–1647, 2020.

[18] M. Yao, C. Chelmis, and D.-S. Zois, "Cyberbullying ends here: Towards robust detection of cyberbullying in social media," in *The World Wide Web Conference*, 2019, pp. 3427–3433.

[19] M. T. Zaman, M. Zaman, F. Shah, and E. Ahmed, *A deep learning-based bengali visual question answering system using contrastive loss*, Available at ResearchGate, Apr. 2024. DOI: 10.13140/RG.2.2.32405.13288.