

## Article

# Exhaustive Study into Machine Learning and Deep Learning Methods for Multilingual Cyberbullying Detection in Bangla and Chittagonian Texts

Tanjim Mahmud <sup>1,2,\*</sup> , Michal Ptaszynski <sup>1,\*</sup>  and Fumito Masui <sup>1</sup> 

<sup>1</sup> Text Information Processing Laboratory, Kitami Institute of Technology, 165 Koen-cho, Kitami City 090-8507, Hokkaido, Japan; f-masui@mail.kitami-it.ac.jp

<sup>2</sup> Department of Computer Science and Engineering, Rangamati Science and Technology University, Rangamati 4500, Bangladesh

\* Correspondence: tanjim\_cse@yahoo.com (T.M.); michal@mail.kitami-it.ac.jp (M.P.)

**Abstract:** Cyberbullying is a serious problem in online communication. It is important to find effective ways to detect cyberbullying content to make online environments safer. In this paper, we investigated the identification of cyberbullying contents from the Bangla and Chittagonian languages, which are both low-resource languages, with the latter being an extremely low-resource language. In the study, we used both traditional baseline machine learning methods, as well as a wide suite of deep learning methods especially focusing on hybrid networks and transformer-based multilingual models. For the data, we collected over 5000 both Bangla and Chittagonian text samples from social media. Krippendorff's alpha and Cohen's kappa were used to measure the reliability of the dataset annotations. Traditional machine learning methods used in this research achieved accuracies ranging from 0.63 to 0.711, with SVM emerging as the top performer. Furthermore, employing ensemble models such as Bagging with 0.70 accuracy, Boosting with 0.69 accuracy, and Voting with 0.72 accuracy yielded promising results. In contrast, deep learning models, notably CNN, achieved accuracies ranging from 0.69 to 0.811, thus outperforming traditional ML approaches, with CNN exhibiting the highest accuracy. We also proposed a series of hybrid network-based models, including BiLSTM+GRU with an accuracy of 0.799, CNN+LSTM with 0.801 accuracy, CNN+BiLSTM with 0.78 accuracy, and CNN+GRU with 0.804 accuracy. Notably, the most complex model, (CNN+LSTM)+BiLSTM, attained an accuracy of 0.82, thus showcasing the efficacy of hybrid architectures. Furthermore, we explored transformer-based models, such as XLM-Roberta with 0.841 accuracy, Bangla BERT with 0.822 accuracy, Multilingual BERT with 0.821 accuracy, BERT with 0.82 accuracy, and Bangla ELECTRA with 0.785 accuracy, which showed significantly enhanced accuracy levels. Our analysis demonstrates that deep learning methods can be highly effective in addressing the pervasive issue of cyberbullying in several different linguistic contexts. We show that transformer models can efficiently circumvent the language dependence problem that plagues conventional transfer learning methods. Our findings suggest that hybrid approaches and transformer-based embeddings can effectively tackle the problem of cyberbullying across online platforms.

**Keywords:** multilingual models; low-resource languages; machine learning; ensemble models; deep learning; hybrid models; transformers models



**Citation:** Mahmud, T.; Ptaszynski, M.; Masui, F. Exhaustive Study into Machine Learning and Deep Learning Methods for Multilingual Cyberbullying Detection in Bangla and Chittagonian Texts. *Electronics* **2024**, *13*, 1677. <https://doi.org/10.3390/electronics13091677>

Academic Editor: Arkaitz Zubiaga

Received: 14 March 2024

Revised: 17 April 2024

Accepted: 24 April 2024

Published: 26 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The rapid expansion of the Internet over the past two decades has ushered in a transformative era in Bangladesh, with the Bangladesh Telecommunication Regulatory Commission (BTRC) reporting a quickly growing user base exceeding 125 million as of November 2022 [1]. This surge highlights the country's remarkable digital trajectory, which has been further bolstered by initiatives like Digital Bangladesh [2], which has significantly expanded Internet

access, especially in Chittagong [3], the nation's second-largest city. In Chittagong, Internet adoption has become synonymous with active participation in social media platforms, thus propelling Bangladesh's online presence, particularly on platforms like Facebook [4]. Within this digital landscape, linguistic diversity thrives, with Chittagonian and Bangla languages emerging as primary modes of communication [5,6].

Chittagonian, spoken by approximately 13 million people, adds to Bangladesh's rich linguistic tapestry [5]. The widespread use of Unicode across communication devices has enabled individuals, including Chittagonian speakers, to freely express themselves in their native tongues [7,8]. Social media platforms like Facebook, imo, WhatsApp, or various blog services have been embraced by the people of Chittagong, thereby fostering an environment conducive to uninhibited self-expression [9–11].

However, this digital revolution is not devoid of challenges. The allure of social media, particularly among youth, has led to excessive usage, potentially bordering on addiction [12]. Such behavior can weaken social interactions and family ties, thereby adversely impacting overall well-being [13]. Moreover, there has been a concerning rise in online abuse, cyberbullying, and the dissemination of hateful content, thereby contributing to a distressing increase in physical violence [14].

In response to these pressing issues, this paper aims to create a multilingual classification model specifically tailored to combat cyberbullying in Bangla and Chittagonian texts. The overarching goal is to leverage technology to foster a safe and respectful online environment for both languages, thus drawing upon Bangladesh's unique linguistic and cultural context for the betterment of society as a whole.

The research framework we applied to achieve this goal encompasses a comprehensive exploration of various machine learning and deep learning techniques [15–18]. The array of ML techniques includes Logistic Regression, Support Vector Machines, Multinomial Naive Bayes, Decision Trees, Random Forest, K-Nearest Neighbors, as well as ensemble methods like Bagging, Boosting, and Voting. Especially, deep learning has played a pivotal role in advancing many important problems in signal processing [19–21], image processing [22–24], and medical diagnosis [25–27]. The deep learning methods applied in this study include Multilayer Perceptron, Simple Recurrent Neural Networks, Long Short-Term Memory, Bidirectional Long Short-Term Memory, Gated Recurrent Unit, Convolutional Neural Networks, as well as hybrid models DL such as BiLSTM+GRU, CNN+LSTM, CNN+BiLSTM, CNN+GRU, and (CNN+LSTM)+BiLSTM, along with Transformer-based models like BERT, Bangla BERT, Bangla ELECTRA [28], Multilingual BERT, and XLM-Roberta.

The subsequent sections elaborate in detail on the process and outcomes of this study, thus offering insights that could bolster strategies for addressing cyberbullying in the Bangla and Chittagonian linguistic domains.

The research objectives of this study were as follows:

1. To curate a comprehensive dataset comprising at minimum five thousand manually collected samples from both Bangla and Chittagonian languages, thereby ensuring balanced representation.
2. To ensure the production of high-quality ground truth data through meticulous manual labeling by human annotators validated using Krippendorff's alpha [29] and Cohen's kappa [30] scores.
3. To conduct an exhaustive analysis starting with baseline methods based on traditional machine learning (ML) models, including Support Vector Machine (SVM), Bagging, Boosting, and Voting, to determine their effectiveness in cyberbullying detection.
4. To explore deep learning (DL) models, particularly Convolutional Neural Network (CNN) architectures, and assess their performance compared to traditional ML approaches.
5. To propose and evaluate hybrid network-based models, such as BiLSTM+GRU, CNN+LSTM, CNN+BiLSTM, and CNN+GRU for cyberbullying detection.

6. To investigate Transformer-based models, including BERT, Bangla BERT, Bangla ELECTRA, Multilingual BERT, and XLM-Roberta, and estimate their impact on enhancing the accuracy levels in cyberbullying detection.

The paper follows a structured approach beginning with a review of the literature on multilingual cyberbullying detection, which is presented in Section 2. Section 3 details the proposed methodology, thereby encompassing dataset acquisition, annotation, and preprocessing, with a specific focus on Chittagonian and Bangla nuances and the application of NLP, ML, and DL techniques. In Section 4, the experimental results and evaluation procedures are detailed, thus offering insights into the effectiveness of the proposed methodologies. Ethical considerations pertinent to the study are deliberated upon in Section 4.9. Finally, the paper concludes in Section 5, where the contributions are summarized, and avenues for future research are suggested.

## 2. Related Work

Cyberbullying detection has gained considerable attention in recent years owing to the widespread use of social media platforms and online communication channels [7]. Researchers have explored various techniques and methodologies to identify and address cyberbullying among various languages and cultures [7,8,31–35]. However, while numerous research efforts have introduced solutions to detect cyberbullying in high-resource languages such as English or Japanese, there is a limited number of studies that have extensively addressed cyberbullying detection in the low-resource languages, such as the Bangla language. Furthermore, to date, there has been no research specifically targeting both Bangla and its dialects and similar languages simultaneously. This section reviews the pertinent literature on cyberbullying detection, multilingual text analysis, and studies specifically focusing on low-resource languages like Arabic [16], Hindi [15,28], Marathi [15], Indonesian [18], and Turkish [36].

Pawar et al. [15] suggested developing a multilingual system to identify cyberbullying in Hindi, Marathi, and English. It used lexicon-based and machine learning classification techniques to determine whether or not the input data represented cyberbullying. They experimented with several prototypes (standalone, collaborative, and cloud-based) to detect cyberbullying in a variety of datasets and languages. The results of their tests demonstrated that the machine learning model outperformed the lexicon-based model in all languages. Furthermore, the outcomes demonstrated that cooperation strategies could raise a system node's accuracy that is not performing at its peak. Ultimately, they discovered that the cloud-based configurations outperformed the local configurations, covering Marathi, Hindi, and English in obtaining high levels of accuracy (0.8106 for the lexicon-based approach and 0.9024 for the LR-based classifier).

Haidar et al. [16] presented a multilingual model for Arabic and English cyberbullying detection with notable SVM F1 scores of 0.927 and NB scores of 0.905. Mahajan et al. [18] investigated the use of deep learning methods to improve the accuracy of detection. There have been positive improvements in the overall performance observed in models that use Bidirectional Long Short-Term Memory (BiLSTM), the Bidirectional Gated Recurrent Unit (Bi-GRU), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM). Furthermore, the use of Global vectors for word Representation (GloVe) embeddings made it easier to translate words into vector representations, which sped up the processing of multilingual data streams. Novel methods have surfaced to augment detection efficacy, including hybrid ensemble techniques. By combining several effective models, like CNN-LSTM, these techniques optimize weights and increase model accuracy through the use of bagging ensemble methods and stacking approaches. Moreover, they outperformed alternatives with at least a 4.44% increase in F1 score.

Michele et al. [37] proposed neural architecture that functioned reasonably well in a number of languages, including Italian, German, and English. A detailed examination of the experimental findings in all three languages was carried out in order to better understand the contributions made by the various system components, such as long

short-term memory, gated recurrent units, and bidirectional long short-term memory. In addition, word embeddings, emotion expressions, n grams, social network-specific features, and emojis were among the features chosen for the system. To conduct such a thorough analysis, they used three publicly available datasets for the identification of hate speech on social media in the languages of English, Italian, and German.

Shukrity et al. [38] dealt with English, Hindi, and Hindi–English (mixed code) datasets. They used features like word vectors and manually assembled dictionaries for aggressive words, sentiment scores, parts of speech, and emojis for the classification task. They discovered that the Gradient Boosting Classifier (GBM), XGBoost Classifier, and Support Vector Machine (SVM) were the best models for the task after experimenting with a variety of machine learning and deep learning models. Therefore, majority voting was done using the output of the three classifiers, which produced F1 scores of 68.13, 54.82, and 55.31 for the English, Hindi, and mixed code datasets, respectively.

Sayar Ghosh et al. [39] examined how challenging it is to identify offensive or hateful content on Twitter. Three categories of harmful content were identified: profane (PRFN), offensive (OFFN), and hate speech (HATE). They first used a Transformer-based text encoder that was multilingual and pretrained it to identify and classify hate speech in a range of languages. When they performed hate speech detection on the provided testing corpora, they received macro F1 scores of 60.70, 53.28, and 49.74 for fine-grained classification and 90.29, 81.87, and 75.40 for English, German, and Hindi, respectively.

El-Alami et al. [40] applied a model based on Bidirectional Encoder Representations from Transformers (BERTs), which is helpful in extracting semantic and contextual information from texts. Their system consisted of three stages: preprocessing, text representation with BERT models, and classification into offensive and nonoffensive categories. To address multilingualism, they looked into a number of techniques, including translation-based and joint multilingual approaches. Whereas all texts were translated into a single universal language prior to classification in the latter scenario, a single classification system was developed in the former scenario for multiple languages. For several tests, they applied a bilingual dataset from the Semisupervised Offensive Language Identification Dataset (SOLID). The translation-based approach combined with Arabic BERT (AraBERT) produced an accuracy and an F1 score of over 91% and 93%, respectively, based on the experimental results.

To detect hate speech on various social media platforms, including Facebook, Instagram, WhatsApp, Twitter, and Facebook, Arijit et al. [41] suggested a Transformer-based model. Datasets with contents in Bengali, English, Italian, and German were used for testing. In the evaluation, based on prepared gold standard datasets, the accuracy rates for the Bengali dataset were 89%, while the English, German, and Italian datasets obtained accuracy rates of 91%, 91%, and 77%, respectively. Their proposed model was able to identify hate speech with greater accuracy than both baseline and state-of-the-art models.

For the purpose of identifying cyberbullying in texts with mixed Hindi–English (Hinglish) and English codes, Malte et al. [28] proposed a deep learning generative method. Bidirectional transformer-based BERT architecture was used in this technique. The proposed architecture yielded state-of-the-art results on the mixed code Hindi dataset and satisfying results on the English task leaderboard in the TRAC-1 standard aggression identification task. The results showed an improved alternative to the existing approaches because they could be obtained without requiring multiple models or ensemble-based methods. Deep learning-based models that performed well on multilingual text were considered to be crucial to addressing the problem of cyberbullying because they could handle a wider range of inputs.

Through transfer learning and crosslingual contextual word embeddings, Ranasinghe et al. [36] produced predictions in low-resource languages using English data at their disposal. They reported results of a 0.8415 F1 macro for Bengali, a 0.8568 F1 macro for Hindi, and a 0.7513 F1 macro for Spanish after projecting predictions onto comparable data in those three languages. Lastly, they showed that their method outperformed the top

systems that had been submitted in recently shared tasks on these three languages, thus proving the usefulness of transfer learning and crosslingual contextual embeddings for similar tasks.

### 3. Data and Methods

#### 3.1. Data Collection

The data collection for this study was conducted manually from April to June 2022. A total of 5120 Bangla and 5146 Chittagonian samples were obtained from a diverse set of online platforms, including social media sites, forums, and comment sections of various websites.

Figure 1 shows the steps that encompassed the comprehensive process involved in both obtaining and managing the data to ensure their effectiveness and accessibility for further investigation. These datasets comprise text samples in both Chittagonian and Bangla languages, thus encompassing various types of content such as posts and comments (refer to Figure 2). Figure 3 illustrates a selection of data samples.

To ensure a comprehensive representation, we systematically gathered political, religious, and COVID-19-related content, as these topics often provoke intense debates and potential cyberbullying incidents. Furthermore, a wide variety of sources comprise the majority of the dataset, including blogs, YouTube channels, news platforms, Facebook pages, and various Facebook groups like Chatgaiya Express [42] and Chatgaiya Tourist Gang [43].

The dataset encompasses a range of cyberbullying instances, thus facilitating the training and testing of deep multilingual models. Our objective was to establish a robust foundation for evaluating the effectiveness of cyberbullying detection methods in both languages through this extensive data collection process.

The collected data served as the cornerstone of our comparative analysis, thus enabling a deeper understanding of cyberbullying trends and the performance of the developed detection models. For a summary of both Bangla and Chittagonian datasets, please refer to Table 1.

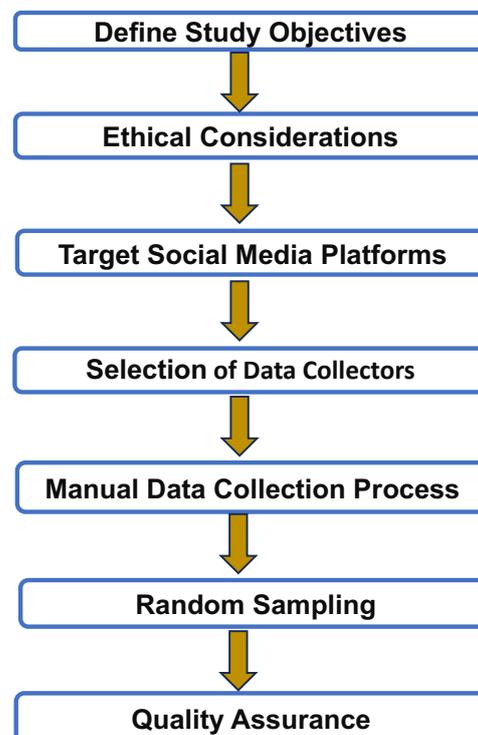
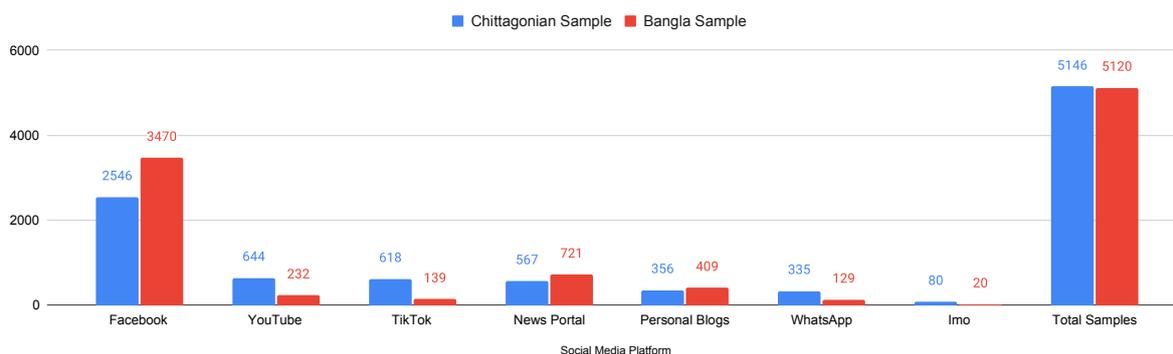


Figure 1. Procedures implemented to gather and arrange data.

**Table 1.** Dataset statistics.

Specifications	Bangla	Chittagonian
Number of Comments/Posts	5120	5146
Number of Words	72,327	44,314
Vocabulary Size	15,909	15,200
Number of Sentences	5516	5448
Average Number of Words per Sentence	13.11	8.13
Highest number of words in a sentence	96	51
Lowest number of words in a sentence	1	1



**Figure 2.** Sources of data.

Chittagonian Sample	Bangla Sample	Source	Label
ভোরে কুতাইও বিয়া গইরতু নু। English: A dog will not marry you.	জাতীয় পতিভা English: National Prostitute	Facebook	1 (Bully)
আরা চউগামের মানুষ দেরা English: We people of Chittagong are the best	শেষের টা দারুণ ছিল English: The last one was great	Facebook	0 (not-bully)
বাংলাদেশত নতুন নতুন বেইশ্যা দেহা যার। English: New prostitutes are appearing in Bangladesh.	মাসুদ আর একটা মাদারচোদ English: Masud is a complete motherfucker	YouTube	1 (Bully)
শুভ জন্মদিন প্রিয় গ্রুপ,, বউত বউত English: Happy Birthday dear Team,, Many Many"	গতকাল খেলা দেখে আমার অনেকভালো লেগেছে English: I really enjoyed watching the game yesterday	TikTok	0 (not-bully)

**Figure 3.** Data samples.

### 3.2. Data Annotation

#### 3.2.1. Annotator Selection

To ensure optimal data quality, the effectiveness of any cyberbullying detection system hinges not only on the quantity but primarily on the quality of annotated data. To achieve this, we meticulously designed a systematic data annotation process.

We leveraged the expertise and fluency of three native speakers proficient in both Chittagonian and Bangla. These annotators, consisting of two females and one male, possess advanced degrees in linguistics and demonstrate a deep understanding of the intricacies of both languages. The inclusion of gender diversity among annotators was deliberate, thereby aiming to capture a diverse range of linguistic nuances that may vary between male and female speakers.

#### 3.2.2. Data Annotation Guidelines

In the data annotation guidelines of this study, annotators were instructed to categorize instances of cyberbullying within Bangla and Chittagonian texts based on predefined criteria. Annotations primarily focused on identifying various forms of cyberbullying, including but not limited to offensive language, harassment, threats, and derogatory remarks. Annotators were encouraged to consider cultural nuances and language-specific

expressions prevalent in Bangla and Chittagonian texts while evaluating the severity and context of cyberbullying instances. Additionally, guidelines emphasized the importance of consistency and accuracy in annotation practices to ensure the reliability and validity of the annotated dataset for training and evaluating transformer models aimed at multilingual cyberbullying detection.

The data annotation guidelines applied in this study were structured as follows:

- a. Furnish annotators with explicit definitions and examples for clarity.
- b. Label each post as either 'bullying' or 'not-bullying' in accordance with predetermined definitions.
- c. Take cultural nuances into account during the labeling process.
- d. Conduct regular meetings with annotators to ensure consensus.
- e. Address language variations and slang to maintain accuracy in annotations.

### 3.2.3. Data Annotation Results

The data annotation methods employed in the study on multilingual cyberbullying detection in Bangla and Chittagonian languages, depicted in Figure 4, yielded a dataset with the following distribution:

In Bangla, the dataset comprised 2545 instances labeled as bullying, 2555 instances labeled as not bullying, 11 instances of conflicts (after discussions with annotators, these were excluded and discarded from the final dataset), and 9 instances marked as ignored.

Conversely, in the Chittagonian language, the distribution included 2610 instances of bullying, 2490 instances of not bullying, 25 instances of conflicts, and 21 instances labeled as ignored (refer to Figure 5).

These annotations established a robust foundation for training and evaluating multilingual cyberbullying detection models, thus providing insights into the prevalence of various content categories in online interactions across both languages.

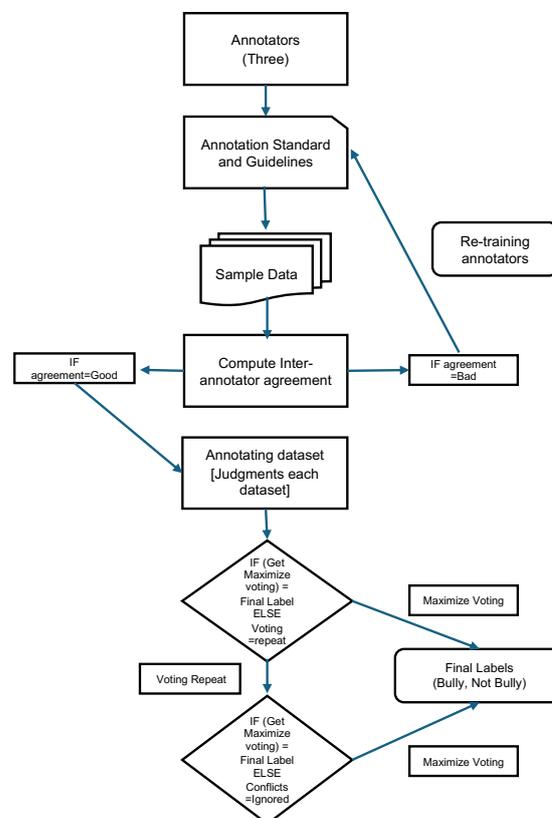
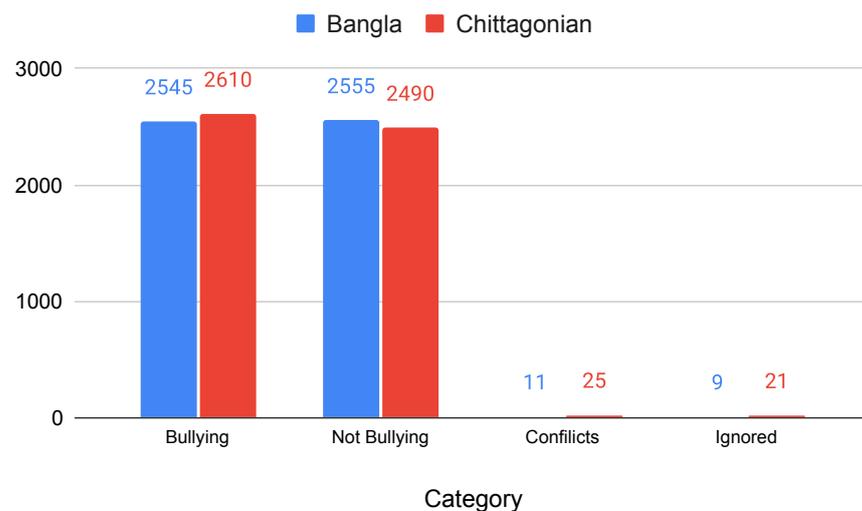


Figure 4. Data annotation process.



**Figure 5.** Summary of data annotation.

### 3.2.4. Quality Evaluation Of Data And Annotation

We rigorously assessed the quality of data and annotations through inter-rater reliability measures. Specifically, for the Chittagonian language, the annotations achieved high Krippendorff's alpha (See Box 1) [29] values ranging from 0.9342 to 0.9641 and Cohen's kappa (See Box 2) [30] values ranging from 0.9281 to 0.9484 among annotator pairs, thereby indicating a commendable level of agreement in labeling cyberbullying instances (refer to Figure 6). This consistency in annotations underscores the dataset's high reliability and suitability for training cyberbullying detection models, thereby enhancing the credibility and effectiveness of the research presented.

#### Box 1. Krippendorff's alpha calculation.

The formula for calculating Krippendorff's alpha involves four steps:

1. **Compute Observed Agreement ( $P_o$ ):** Tally agreements and disagreements between annotator pairs. It is calculated by

$$P_o = \frac{\sum \text{Agreements}}{\sum (\text{Agreements} + \text{Disagreements})}$$

2. **Compute Expected Agreement ( $P_e$ ):** Calculate the proportion of expected agreement under the assumption of random labeling.
3. **Calculate Disagreement Matrix ( $D_o$ ):** This matrix captures the differences between annotator responses for each item in the dataset.
4. **Calculate Alpha Value:** Use observed and expected agreement to compute Krippendorff's alpha using the following formula:

$$\alpha = 1 - \frac{D_o}{P_e}$$

(where  $D_o$  is the mean of the squared differences between observed and expected disagreement, and  $P_e$  is the sum of squared differences between expected agreement and the product of observed agreement and its complement.)

Similarly, for the Bangla language, Krippendorff's alpha and Cohen's kappa values consistently surpassed 0.8 for all annotator pairs, thereby indicating substantial to near-perfect agreement among annotators (refer to Figure 7). These metrics reflect a meticulous annotation process, clear guidelines, and a strong foundation for data reliability. The high inter-rater reliability underscores the accuracy and consistency of the data annotations,

thereby substantiating their efficacy for training and evaluating deep learning models targeting cyberbullying detection in both Bangla and Chittagonian languages.

**Box 2.** Cohen’s kappa calculation.

The formula for calculating Cohen’s kappa involves four steps:

1. **Construct Contingency Table:** Tabulate the agreement and disagreement frequencies between annotator pairs into a contingency table.
2. **Calculate Proportional Agreement:** Determine the observed proportion of agreement ( $P_o$ ) by summing the diagonal elements (representing agreement) of the contingency table and dividing by the total number of observations.

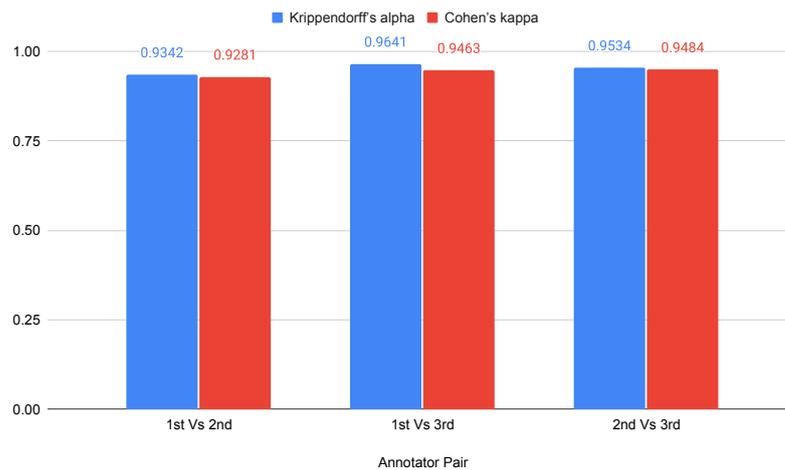
$$P_o = \frac{\text{Agreements}}{\text{Total Observations}}$$

3. **Calculate Chance Agreement:** Compute the proportion of agreement expected by chance ( $P_e$ ) based on the marginal probabilities of each label in the contingency table.

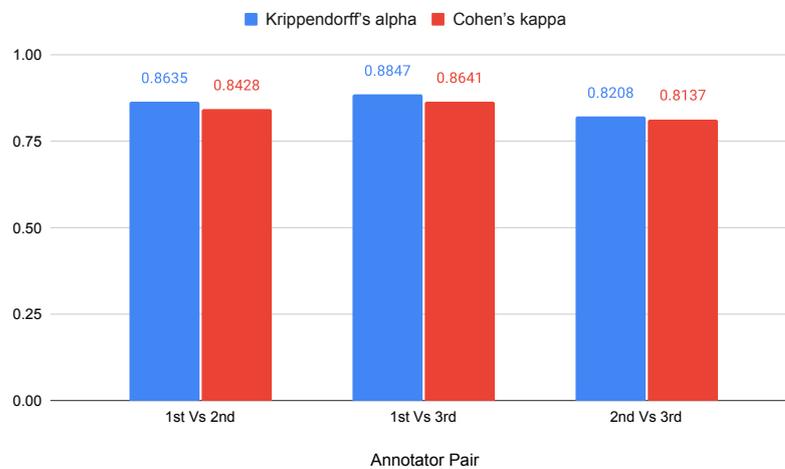
$$P_e = \frac{\sum(\text{Row Total}) \times \sum(\text{Column Total})}{\text{Total Observations}^2}$$

4. **Compute Kappa Value:** Calculate Cohen’s kappa ( $\kappa$ ) using the following formula:

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$



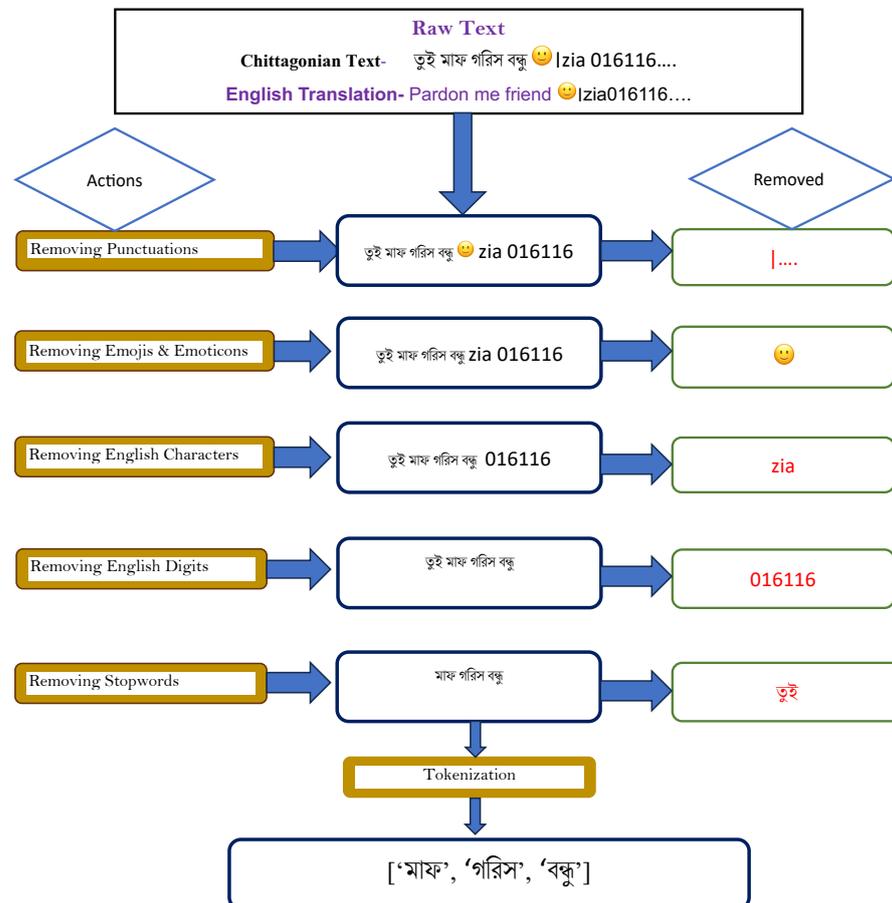
**Figure 6.** Inter-rater reliability results for Chittagonian language.



**Figure 7.** Inter-rater reliability results for Bangla language.

### 3.3. Data Preprocessing

The data preprocessing procedure for the multilingual cyberbullying detection in Bangla and Chittagonian languages encompassed a series of sequential steps as shown in Figure 8. Python's NLTK library [44,45] was used for the elimination of stop words, while Matplotlib and Seaborn were utilized for data visualization. Additionally, the Pandas library facilitated the preparation of the data for the subsequent stages of analysis.



**Figure 8.** Data preprocessing steps.

#### 3.3.1. Removing Punctuations

In this study, the applied text data comprises various punctuation and special characters, which serve diverse grammatical functions such as denoting pauses, emphasis, and other structural aspects in written language. These characters encompass periods, commas, question marks, exclamation marks, hyphens, parentheses, quotation marks, and other nonalphanumeric symbols. While some punctuation may not significantly alter the meaning of a sentence, their presence can impact text classification tasks, including automatic cyberbullying detection, as demonstrated by Mahmud et al. [7] in previous research. Their findings suggest that removing punctuation can enhance text classification accuracy, particularly when employing traditional machine learning algorithms. Hence, we adopted a similar approach and removed all punctuation from the text during the analysis. The list of punctuation considered for removal included characters such as periods, commas, quotation marks, exclamation marks, hashtags, parentheses, and various other symbols commonly found in written text found on the Internet.

### 3.3.2. Removing Emoji and Emoticons

In written communication, emoticons and emojis serve as graphical representations of emotions, expressions, or symbols, thus offering users a means to convey feelings and reactions in online interactions. These elements play a significant role in adding context and enhancing the emotional tone of text-based conversations. When handling such information in text analysis, two approaches are commonly adopted: either removing emoticons and emojis entirely from the text or substituting them with corresponding textual representations.

While research has demonstrated that both emoticons [46] and emojis [47] can contribute to text classification tasks in certain scenarios [47], we chose not to consider emojis and emoticons during the classification process. As a result, we opted to exclude them from our analysis to maintain a focus on textual content without graphical representations of emotions and expressions.

### 3.3.3. Removing English Characters

In multilingual text analysis, ensuring the purity and integrity of language data is paramount for the precise performance of models developed specifically for the specific linguistic context. Therefore, removing English characters from Bangla and Chittagonian text becomes imperative. Bangla and Chittagonian, as distinct languages, possess their unique scripts and character sets, which contribute to their linguistic identity. Mixing English characters with Bangla or Chittagonian text introduces noise and inconsistencies, thereby potentially impeding the effectiveness of the multilingual models trained on these datasets. Standardizing preprocessing procedures by eliminating extraneous characters aids in maintaining data integrity and ensuring that models learn the specific linguistic patterns and nuances inherent to Bangla and Chittagonian. This preprocessing step not only enhances model robustness but also fosters language consistency, thereby enabling the model to focus solely on the linguistic characteristics of the target languages. By avoiding confusion during training and inference, the models can more accurately process and classify text, thereby advancing the effectiveness of multilingual text analysis throughout a variety of language corpora.

### 3.3.4. Removing English Digits

After closely examining the samples in the dataset, we found that some of them contained digits and numbers with no apparent semantic meaning. Phone numbers, currencies, and percentages are examples of numerical entities that are commonly identified and categorized using Named Entity Recognition (NER) tools [48]. However, there is currently no NER tool available for Chittagonian text. Therefore, we decided to eliminate the digits and numbers from the dataset in order to get around this restriction and streamline our initial experiment. Nevertheless, we acknowledge that precise manipulation of numerical entities is necessary for subsequent investigations. For this reason, we plan to create a Chittagonian-only NER tool in the future.

### 3.3.5. Removing Stopwords

Words that are frequently used in a language but typically do not significantly change the meaning of a sentence are known as stopwords. When processing text data for tasks involving natural language processing, stopwords are often omitted. In the Chittagonian language, such words include, for example, 'কি(What)', 'তুই(You)', 'হি(He)', 'তুই(You)', 'তোর(Yours)', 'তোয়ারে(You)' etc. In many natural language processing tasks, like text classification, they are frequently eliminated in order to decrease the dataset dimensionality [34]. We thus decided to eliminate stopwords from the dataset in order to reduce the dimensionality of the text data and boost the efficacy and efficiency of the subsequent analyses.

### 3.3.6. Tokenization

In preparation for cyberbullying detection analysis in Bangla and Chittagonian languages, we conducted tokenization, which involves dividing the text into tokens or individual words. This process enables the segmentation of text into meaningful linguistic units, such as words or subwords, thereby facilitating further analysis. Our approach to tokenization aligns with previous research on the Chittagonian language, which also emphasized the importance of breaking down text into linguistic units [7].

### 3.4. Feature Extraction

In this study, for feature extraction, we utilized Term Frequency–Inverse Document Frequency (TF-IDF) [49–51] for classic machine learning and Word2Vec embeddings [52,53] for deep learning algorithms. TF-IDF functions as a conventional feature extraction method for ML, thus capturing the significance of words in numerical vectors. Conversely, DL employs Word2Vec embeddings to represent semantic meaning and encapsulate word relationships in vector space. This strategy extracted language nuances specific to Bangla and Chittagonian languages, thereby enhancing the efficacy of cyberbullying detection.

#### 3.4.1. TF-IDF

The TF-IDF method transforms text data into numerical vectors by calculating the importance of terms based on their frequency in each document and rarity across the entire dataset [49,54]. Each document is represented as a vector, where each element corresponds to the TF-IDF score of a term. This vectorization process enables the use of machine learning algorithms for tasks such as classification and clustering. By capturing the unique linguistic characteristics of Bangla and Chittagonian languages, the TF-IDF method facilitates the development of effective multilingual models for tasks like cyberbullying detection. Through the integration of TF-IDF, ML models can recognize patterns and extract relevant word features from text data, which allows them to achieve high accuracy with competitive applicability in a variety of linguistic circumstances [31,32].

#### 3.4.2. Word2Vec

In multilingual models tailored for Bangla and Chittagonian languages, Word2Vec embeddings serve as crucial components for encoding semantic information and capturing intricate linguistic relationships [53,55]. These embeddings transform words into dense vector representations, where each word's position in the vector space reflects its semantic context and proximity to other words. By leveraging Word2Vec embeddings, multilingual models can effectively discern subtle nuances specific to Bangla and Chittagonian, thus facilitating more accurate and contextually relevant cyberbullying detection [34].

### 3.5. Classification Algorithms

In this section, we present a diverse array of methods aimed at addressing the challenge of multilingual cyberbullying detection in the Bangla and Chittagonian languages. Our investigation begins with traditional machine learning algorithms, including Support Vector Machine, Logistic Regression, Decision Tree, Random Forest, and K-Nearest Neighbors. These algorithms serve as foundational techniques in our exploration.

Furthermore, we employ ensemble models such as Bagging, Boosting, and Voting classifiers to enhance the robustness and accuracy of our detection framework. These ensemble methods leverage the collective decision making of multiple composing models to improve overall classification performance.

Transitioning to deep learning architectures, we further apply the efficacy of various neural networks, such as Multilayer Perceptron, Recurrent Neural Networks encompassing SimpleRNN and Long Short-Term Memory, Gated Recurrent Units, and Convolutional Neural Networks. Each architecture offers unique advantages in capturing complex patterns and features present in cyberbullying text data.

Moreover, we evaluate the performance of hybrid architectures, namely BiLSTM+GRU, CNN+LSTM, CNN+GRU, CNN+BiLSTM, and (CNN+LSTM)+BiLSTM. These hybrid models combine the strengths of multiple architectures, thus harnessing the complementary nature of different neural network components to improve classification accuracy and generalization.

In addition to exploring neural network architectures, we employ Transformer-based models including XLM-Roberta, Bangla BERT, Multilingual BERT, BERT, and Bangla ELECTRA. These state-of-the-art models leverage pretrained representations to capture contextual information and semantic relationships, thereby enhancing the detection accuracy across multilingual text data.

Overall, our classification framework integrates a comprehensive range of methods, which we test on the nuances of Bangla and Chittagonian languages. Through rigorous evaluation and comparison, we aim to discern the most effective approaches for multilingual cyberbullying detection, thereby providing insights into the advancement of cyberbullying detection technology.

### 3.5.1. Classic Machine Learning Algorithms

#### Logistic Regression

In our multilingual model for Bangla and Chittagonian languages, Logistic Regression (LR) serves as a fundamental baseline classification algorithm [51]. LR leverages probabilistic estimation to categorize input text into cyberbullying or noncyberbullying categories. Its simplicity and efficiency make it an attractive choice for multilingual text classification tasks. By analyzing the linguistic characteristics specific to Bangla and Chittagonian, LR effectively learns discriminative patterns to differentiate between cyberbullying and noncyberbullying content. Through comprehensive evaluation and fine-tuning, LR contributes to the robustness and accuracy of our multilingual cyberbullying detection framework.

#### Support Vector Machine

Support Vector Machine (SVM) serves as a strong baseline method for constructing a multilingual cyberbullying detection model [16]. SVM is adept at separating data points into different classes by finding the optimal hyperplane that maximizes the margin between classes. In our study, SVM is applied to the multilingual dataset comprising both Bangla and Chittagonian text samples. By leveraging SVM, we aim to discern distinctive patterns and features indicative of cyberbullying behavior in a variety of linguistic settings. The versatility of SVM allows it to handle high-dimensional data efficiently, thus making it suitable for processing text data with varying complexities. Thorough assessment, we assess the effectiveness of SVM in classifying instances as cyberbullying, thereby contributing to the advancement of multilingual cyberbullying detection methodology.

#### Multinomial Naive Bayes

In our multilingual model for cyberbullying detection, we also leverage Multinomial Naive Bayes (MNB), a versatile algorithm known for its simplicity and effectiveness in text classification tasks. MNB operates based on the assumption of independence between features, thereby making it particularly suitable for processing text data. By calculating the likelihood of each class given a set of input features, MNB efficiently assigns probabilities to different categories, thus facilitating accurate classification. Despite its simplicity, MNB has demonstrated robust performance in various natural language processing applications, including sentiment analysis [16] and spam detection [56,57].

#### Decision Trees

Decision Trees (DTs) serve as another baseline for our multilingual cyberbullying detection study. Decision Trees recursively partition the feature space based on attribute values, thereby enabling the classification of text data into distinct categories [58]. In our study, DTs offer an interpretable and transparent approach to extracting linguistic patterns

indicative of cyberbullying across Bangla and Chittagonian texts. By recursively splitting the dataset into homogeneous subsets, DTs effectively capture the discriminatory power of various features present in the text. We carefully evaluate DTs' performance in comparison to other classification techniques in order to determine how effective it is at detecting multilingual cyberbullying.

#### K-Nearest Neighbors

K-Nearest Neighbors (kNNs) is an instance-based machine learning algorithm popularly used as a baseline, which has also been used in previous studies on multilingual cyberbullying detection [59]. It operates based on the principle of similarity, where instances are classified by the majority class of their K-Nearest Neighbors. The simplicity of KNNs and its intuitive approach make it an attractive choice for classifying text data in linguistic contexts. However, its performance heavily depends on the choice of distance metric and the value of  $k$ , which may require careful tuning. Despite its straightforward implementation, kNNs may face challenges in high-dimensional spaces and with large datasets, thereby potentially affecting their computational efficiency and classification accuracy. Nonetheless, when appropriately applied and optimized, kNNs can serve as a valuable tool in the multilingual cyberbullying detection pipeline.

#### 3.5.2. Ensemble Methods

##### Random Forest

Random Forest (RF) is an ensemble learning technique utilized in our multilingual offensive text detection model [60]. It operates by constructing multiple decision trees during training and outputs the mode of the classes as the prediction of individual trees. RF is well-suited for handling complex data and is particularly effective in capturing nonlinear relationships between features. In our study, RF serves as one of the traditional machine learning algorithms employed alongside other methods to enhance classification accuracy. Through assessment, we assess the performance of RF in discerning cyberbullying content across Bangla and Chittagonian texts, thereby contributing valuable insights to the field of multilingual cyberbullying detection.

##### Bagging

In our study on multilingual cyberbullying detection, we employed Ensemble Bagging with Random Forest (RF) as a robust classification technique. Bagging, or Bootstrap Aggregating, is a powerful ensemble method [61] that combines multiple base learners to improve overall model performance. By training multiple RF classifiers on different subsets of the dataset and aggregating their predictions, Ensemble Bagging reduces variance and enhances model stability. This approach allows us to mitigate overfitting and capture diverse patterns present in cyberbullying text data.

##### Boosting

In the Boosting approach for cyberbullying detection, we utilize three powerful classifiers: AdaBoost Classifier, XGB Classifier, and Gradient Boosting Classifier. These classifiers are renowned for their ability to leverage the collective strength of multiple weak learners to improve overall prediction accuracy [62]. AdaBoost Classifier adapts by focusing more on misclassified instances, while XGB Classifier implements gradient boosting to optimize performance iteratively. Similarly, Gradient Boosting Classifier constructs a sequence of decision trees to progressively correct errors, thus resulting in enhanced classification performance. By combining the capabilities of these classifiers, our ensemble boosting method aims to achieve satisfactory cyberbullying detection accuracy across Bangla and Chittagonian texts.

### Voting Approach

In our ensemble voting approach, we combine the predictive power of Random Forest, Support Vector Machine, and Gradient Boosting Machine to create a robust and diverse classification ensemble. Each base model, RF, SVM, and GBM, brings its unique strengths in handling complex patterns and relationships within the multilingual cyberbullying detection task. The ensemble voting mechanism aggregates the individual predictions from these models, thereby allowing them to collectively contribute to the final decision-making process [63]. By leveraging the complementary nature of these diverse algorithms, our ensemble voting strategy aims to enhance overall accuracy and robustness, providing a more reliable and effective solution for multilingual cyberbullying detection in Bangla and Chittagonian texts.

### 3.5.3. Neural Network-Based Methods

#### Multilayer Perceptron

Multilayer Perceptron (MLP) is a fundamental neural network architecture widely used in machine learning tasks [64]. It consists of multiple layers of nodes, including an input layer, one or more hidden layers, and an output layer. MLP is characterized by its ability to learn complex patterns and nonlinear relationships within data. By employing activation functions and backpropagation algorithms, MLP can effectively model intricate structures in cyberbullying text data. Its versatility and scalability make MLP a popular choice for various classification tasks, including cyberbullying detection in multilingual contexts.

#### Simple Recurrent Neural Network

The SimpleRNN, or Simple Recurrent Neural Network, is a basic form of recurrent neural network architecture designed to process sequential data by maintaining a hidden state. It operates by taking input at each time step and producing an output while retaining information from previous time steps [65]. Despite its simplicity, SimpleRNNs suffer from the vanishing gradient problem, thereby limiting their ability to capture long-range dependencies in sequences. Consequently, they are less effective in handling long-term dependencies compared to more advanced recurrent architectures like LSTM and GRU. However, SimpleRNNs remain valuable for tasks requiring short-term memory and basic sequence processing.

#### Long Short-Term Memory Network

Long Short-Term Memory (LSTM) is a type of recurrent neural network architecture known for its ability to capture long-term dependencies in sequential data [66]. It achieves this by utilizing a memory cell that can maintain information over extended time periods. The architecture consists of input, output, and forget gates, which regulate the flow of information within the network. LSTM networks are particularly effective for tasks involving sequential data processing, such as natural language processing and time series prediction. Due to its unique architecture, LSTM has been widely adopted in various applications, including language modeling, sentiment analysis, and speech recognition [67]. Its ability to mitigate the vanishing gradient problem makes it well suited for modeling long sequences and handling complex temporal relationships in data.

#### Bidirectional Long Short-Term Memory Network

Bidirectional Long Short-Term Memory (BiLSTM) is a type of recurrent neural network architecture that is adept at capturing long-range dependencies in sequential data [68]. It consists of two LSTM layers, with each processing input sequences in forward and backward directions simultaneously. By leveraging both past and future context, BiLSTM effectively captures contextual information and semantic dependencies in text data. This architecture is particularly well suited for tasks involving natural language processing, including sentiment analysis, named entity recognition [69], and sequence labeling. BiLSTM has gained popularity due to its ability to model complex temporal relationships

and achieve high accuracy in various sequence-based tasks. In the context of cyberbullying detection, BiLSTM offers a promising approach for capturing linguistic nuances and identifying subtle patterns indicative of cyberbullying behavior.

#### Gated Recurrent Units

Gated Recurrent Units (GRUs) are a type of recurrent neural network architecture widely used in natural language processing tasks [70]. GRU models are adept at capturing long-term dependencies in sequential data while mitigating the vanishing gradient problem encountered in traditional RNNs. Unlike traditional RNNs, GRUs employ gating mechanisms to regulate the flow of information within the network, thereby allowing for more efficient information propagation across time steps. This gating mechanism enables GRUs to retain relevant information while discarding irrelevant information, thus enhancing their ability to model sequential data effectively. GRUs have demonstrated promising performance in various text-related tasks, including sentiment analysis, machine translation, and sequence labeling, thus making them a valuable component in deep learning architectures for cyberbullying detection.

#### Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are powerful deep learning architectures commonly used for various tasks including image recognition and natural language processing [34,71,72]. In the context of text classification, CNNs excel at capturing local patterns and features within sequential data such as sentences or documents. They utilize convolutional layers to apply filters over input text, thereby extracting relevant features at different levels of abstraction. CNNs are particularly effective for cyberbullying detection tasks, as they can automatically learn hierarchical representations of textual data, thereby discerning important linguistic patterns associated with cyberbullying behavior. By leveraging the inherent parallelism and parameter sharing of convolutional operations, CNNs can efficiently process large volumes of text data, thus making them suitable for real-time cyberbullying detection systems.

#### 3.5.4. Hybrid Network Models

##### BiLSTM+GRU

The hybrid architecture of BiLSTM+GRU represents a framework for multilingual cyberbullying detection. By combining Bidirectional Long Short-Term Memory and the Gated Recurrent Unit, this model effectively captures both forward and backward contextual information within sequences of text data. The BiLSTM component facilitates bidirectional processing, thereby enabling the model to learn intricate patterns and dependencies in the input text. Meanwhile, the GRU module introduces gating mechanisms that enhance the model's ability to retain relevant information and mitigate vanishing gradient problems. Together, the BiLSTM+GRU architecture offers a robust solution for multilingual cyberbullying detection, thus leveraging the strengths of both LSTM and GRU networks to achieve high classification accuracy in various linguistic contexts.

##### CNN+LSTM

The hybrid CNN+LSTM architecture combines the strengths of Convolutional Neural Networks and Long Short-Term Memory networks for multilingual cyberbullying detection. CNNs excel at extracting spatial features and patterns from textual data, while LSTMs are proficient at capturing sequential dependencies and long-term context. By integrating CNNs and LSTMs, the model can effectively capture both local and global contextual information in multilingual text, thereby enhancing its ability to detect cyberbullying throughout various linguistic contexts. This hybrid architecture offers a promising approach to multilingual cyberbullying detection by leveraging the complementary strengths of CNNs and LSTMs in text analysis.

### CNN+GRU

The hybrid CNN+GRU architecture offers a powerful approach to multilingual cyberbullying detection. By combining the convolutional layers of CNN with the recurrent memory units of the GRU, the model can effectively capture both spatial features and temporal dependencies within the text data. This integration enables the network to learn intricate patterns and nuances present in Bangla and Chittagonian languages, thereby enhancing its ability to discern cyberbullying content accurately.

### CNN+BiLSTM

The hybrid architecture of CNN+BiLSTM combines the strengths of Convolutional Neural Networks and Bidirectional Long Short-Term Memory networks, thus offering a robust solution for multilingual cyberbullying detection. CNNs excel at capturing local patterns and features from text data through convolutional filters, while BiLSTMs effectively model long-range dependencies and sequential information. By integrating these two architectures, the hybrid model leverages both local and global contextual information, thereby enabling it to capture diverse linguistic nuances present in Bangla and Chittagonian languages. Through its synergistic combination of CNN and BiLSTM layers, the hybrid model enhances the accuracy and robustness of multilingual cyberbullying detection systems.

### (CNN+LSTM)+BiLSTM

The hybrid architecture (CNN+LSTM)+BiLSTM is a powerful model designed for multilingual cyberbullying detection. By combining the convolutional neural network for feature extraction with the long short-term memory and bidirectional LSTM layers for sequence modeling, this hybrid approach enables the model to effectively learn complex patterns and relationships present in cyberbullying content. Through its innovative design, the (CNN+LSTM)+BiLSTM architecture offers enhanced performance and robustness in multilingual cyberbullying detection tasks.

### 3.5.5. Transformer Models

#### BERT

Natural language processing tasks have been recently transformed by potent Transformer-based language models. One of the first of such models was BERT, or Bidirectional Encoder Representations from Transformers [73]. BERT is capable of capturing contextual information and semantic relationships within sentences because of its extensive pretraining on text data. Because it is also bidirectional, it can encode each word in a sentence by taking into account both words before and after, thus producing rich contextual embeddings. By fine-tuning BERT for particular downstream tasks [74], like cyberbullying detection, it can achieve high performance by adapting its learned representations to the specifics of the target domain.

#### Bangla BERT

Specialized for the Bangla language, Bangla BERT is a version of the BERT model [75]. It makes use of Transformer-based architecture to extract semantic subtleties and contextual information from Bangla text. Bangla BERT is capable of efficiently processing intricate linguistic patterns in Bangla thanks to pretraining on sizable corpora of text data and fine-tuning on datasets tailored to particular tasks. Promising results have been obtained from its deployment in natural language processing tasks, such as cyberbullying detection, thereby indicating its usefulness in a range of Bangla text analysis applications.

#### Bangla ELECTRA

Bangla ELECTRA is a cutting-edge Transformer-based model created especially for tasks involving the natural language processing of Bangla [76]. It utilizes a pretraining approach that focuses on substituting probable alternatives for input tokens and training a discriminator model to discern between authentic and substituted tokens, thus making use

of the ELECTRA architecture. This method improves model performance and efficiency by capturing semantic subtleties and contextual information found in Bangla text data. The creation of language models specifically suited for Bangla-specific natural language processing applications has advanced significantly with the release of Bangla ELECTRA.

#### Multilingual BERT

A Transformer-based model called Multilingual BERT (or mBERT) was created to efficiently handle tasks involving multilingual text processing. It uses contextual embeddings that have already been trained to capture semantic information from multiple languages at once [77]. Through exposure to a variety of linguistic data, Multilingual BERT has the ability to process and convey the subtleties of various languages, including Bangla. Because of its strong architecture, it can handle a wide range of natural language processing tasks, such as the detection of cyberbullying [35], which makes it an invaluable tool for multilingual text analysis.

#### XLNet-Roberta

For multilingual natural language processing tasks, XLNet-Roberta is a state-of-the-art Transformer-based model [78]. It applies RoBERTa, or a Robustly Optimized BERT Pretraining Approach [79], while also extending its robust architecture to multiple languages. XLNet-Roberta improves mBERT's ability to manage multilingual text data efficiently. Through the utilization of pretrained representations in various languages, XLNet-Roberta is able to accurately grasp complex linguistic subtleties and semantic relationships. The model's sophisticated attention mechanisms allow it to produce high-quality embeddings and contextualize input text, which makes it an effective tool for multilingual cyberbullying detection tasks [35].

#### 3.6. Evaluation Metrics

Analyzing a model's performance using test data is known as model evaluation. The following assessment metrics were applied in this study [16,18,80]: *Precision*, *Recall*, *F1 Score*, and *Accuracy*. These metrics are computed based on the counts of *True Positives (TPs)*, *False Positives (FPs)*, *True Negatives (TNs)*, and *False Negatives (FNs)*. The equations for these metrics are given in Equations (1)–(4).

**Accuracy:** Accuracy is a measure of the proportion of correctly classified instances out of the total number of instances in a dataset.

$$Accuracy = \frac{TruePositives + TrueNegatives}{Total\_Number\_of\_Predictions} \quad (1)$$

**Precision:** Precision measures how many out of a total number of predicted positive observations were correct. It reveals how well the model can avoid the positive labeling of negative instances.

$$Precision = \frac{TruePositives}{(TruePositives + FalsePositives)} \quad (2)$$

**Recall:** Recall is calculated by dividing the number of positive observations predicted by the model by the total number of positive observations in the training set. It shows how fully the model captures the positive class.

$$Recall = \frac{TruePositives}{(TruePositives + FalseNegatives)} \quad (3)$$

F1 Score: By combining recall and precision, we obtain the F1 Score. When there is an imbalance in the class distribution, it strikes a good balance between recall and precision.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{(Precision + Recall)} \quad (4)$$

In the context of these equations, we utilize the following:

*TruePositives* represent the count of positive instances correctly identified as positive. *FalsePositives* denote the count of instances predicted as positive when their actual label is negative. *TrueNegatives* stand for the count of observations accurately classified as negative. *FalseNegatives* refer to the count of observations inaccurately classified as negative when they actually belong to the positive class.

## 4. Results and Discussion

### 4.1. Experimental Setup

In this work, we tackled the problem of multilingual cyberbullying detection in Bangla and Chittagonian texts by thoroughly examining a range of Machine Learning (ML), Ensemble, Deep Learning (DL), and Transformer models. Making use of the processing power offered by Google Colab's T4 GPU, we carried out exhaustive tests to assess these models' ability to identify cyberbullying in a variety of language contexts. Each model category's specific training and parameter tuning procedures were included in the experimental setup to provide a methodical and comprehensive evaluation of the models' performance in this crucial task.

#### 4.1.1. Model Parameters for Classic Machine Learning

**Logistic Regression:** Maximum Iterations (max\_iter): 1000

Random State: 42

**Support Vector Machine:**

Kernel: "linear"

Regularization Parameter (C): 1

Random State: 42

**Multinomial Naive Bayes:**

Alpha (Smoothing Parameter): The alpha parameter was set to 1.0.

Random State: 42

**Decision Tree:**

Criterion: "Gini"

Splitter: "Best"

Random State: 42

**Random Forest:**

Number of Estimators (n\_estimators): 100

Random State: 42

**K-Nearest Neighbors:**

Number of Neighbors (n\_neighbors): 5

Random State: 42

#### 4.1.2. Parameter Settings for Ensemble Learning

**Bagging:**

Base Estimator: RandomForestClassifier with random\_state = 42

Number of Estimators (n\_estimators): 209

Random State: 42

**Boosting:**

Estimators: AdaBoostClassifier with random\_state = 42

XGBClassifier with various hyperparameters set to None and random\_state = 42

GradientBoostingClassifier with random\_state = 42

**Voting:**

Estimators: RandomForestClassifier with random\_state = 42

SVC (Support Vector Classifier) with probability = True and random\_state = 42

GradientBoostingClassifier with random\_state = 42

#### 4.1.3. Hyperparameter Settings for Deep Learning

Batch Size:

This was set to 64 for all models except for the hybrid network, where it was reduced to 32.

Epochs: This was set to 5 for most models and extended to 10 for the hybrid network.

Learning Rate: This was set to 0.01, which was chosen to balance stability and efficiency.

Optimizer: Adam optimizer.

Evaluation Metric: Accuracy.

Loss Function: Binary crossentropy.

#### 4.1.4. Hyperparameter Settings for Transformer Model

Hardware: Google Colab's T4 GPU

Transformer Models: XLM-RoBERTa, BERT, Bangla BERT, mBERT, Bangla ELECTRA

Common Parameters: The same hyperparameters were used for all transformer models except BERT and Bangla ELECTRA, which were trained for 10 epochs.

##### **Data Splitting:**

Test Size: 0.2 (20% of the data reserved for testing)

Random State: 42 (for reproducibility)

##### **Tokenizer Settings:**

Model: XLM-RoBERTa, BERT, Bangla BERT, mBERT, Bangla ELECTRA

Padding: True

Truncation: True

Return Tensors: 'tf' (TensorFlow tensors)

##### **Training:**

Optimizer: Adam

Learning Rate:  $1 \times 10^{-5}$

Loss Function: Sparse Categorical Crossentropy

Metrics: Sparse Categorical Accuracy

Batch Size: 64

Epochs: 5

#### 4.2. Classic Machine Learning for Multilingual Cyberbullying Detection

In our investigation, we conducted a thorough analysis of various machine learning algorithms applied to multilingual cyberbullying detection in Bangla and Chittagonian languages, thus employing an 80%:20% training–testing ratio. We meticulously evaluated algorithm performance using diverse evaluation metrics [16,18], with accuracy as our primary focus, which was supported by the F1 score. Support Vector Machine (SVM) emerged as the most effective model, thus showcasing balanced precision, recall, and F1 scores for both bullying and nonbullying instances, with an accuracy of 0.711. Multinomial Naive Bayes (MNB) closely followed SVM, thus demonstrating similar precision and recall rates across both classes, with an accuracy of 0.71. Logistic Regression (LR) also exhibited moderately satisfying performance, thus achieving an accuracy of 0.69. Decision Trees (DT) and Random Forests (RF) performed reasonably well, with accuracies of 0.66 and 0.70, respectively. However, K-Nearest Neighbors (KNNs) lagged behind in performance metrics, particularly in recall for nonbullying instances, thus resulting in an accuracy of 0.63 (refer to Table 2).

**Table 2.** Classic machine learning model performance.

Model	Class	Precision	Recall	F1 Score	Accuracy
LR	Not Bullying	0.66	0.73	0.69	0.69
	Bullying	0.73	0.66	0.69	
SVM	Not Bullying	0.70	0.71	0.71	<b>0.711</b>
	Bullying	0.71	0.70	0.71	
MNB	Not Bullying	0.69	0.72	0.70	0.71
	Bullying	0.73	0.70	0.72	
DT	Not Bullying	0.64	0.65	0.64	0.66
	Bullying	0.67	0.66	0.67	
RF	Not Bullying	0.67	0.72	0.69	0.70
	Bullying	0.72	0.67	0.70	
KNN	Not Bullying	0.66	0.46	0.54	0.63
	Bullying	0.61	0.78	0.69	

#### 4.3. Ensemble Learning Multilingual Cyberbullying Detection

In evaluating the performance of the Bagging, Boosting, and Voting ensembles for multilingual cyberbullying detection in Bangla and Chittagonian texts, distinct patterns emerged across various metrics.

The Bagging ensemble model demonstrated a commendable balance between precision, recall, F1 score, and accuracy. Notably, its precision for identifying instances labeled as “Not Bullying” was 0.67, with a corresponding recall of 0.74, thus resulting in an F1 score of 0.70. Similarly, Bagging achieved a precision of 0.73 and a recall of 0.65 for identifying instances of “Bullying”, thus resulting in an F1 score of 0.69. Overall, Bagging attained an accuracy of 0.70 for “Not Bullying” and 0.69 for “Bullying” instances (refer to Table 3).

The Boosting ensemble method exhibited similar trends to Bagging, albeit with slightly higher recall rates for the “Not Bullying” class. It achieved a precision of 0.65 and a recall of 0.78 for “Not Bullying”, thus leading to an F1 score of 0.71 and an accuracy of 0.69. For “Bullying” instances, Boosting achieved a precision of 0.75 and a recall of 0.60, thus resulting in an F1 score of 0.67 and an accuracy of 0.67.

In contrast, the Voting ensemble consistently outperformed Bagging and Boosting across all evaluation metrics. With precision, recall, F1 score, and accuracy of 0.69, 0.73, 0.71, and 0.72, respectively, for “Not Bullying” instances and 0.74, 0.70, 0.72, and 0.72, respectively, for “Bullying” instances, the Voting ensemble demonstrated superior performance. These findings highlight how well the Voting ensemble performs when a variety of classifiers are combined, which makes it the best option for precise cyberbullying detection in multilingual environments.

**Table 3.** Ensemble learning model performance.

Model	Class	Precision	Recall	F1 Score	Accuracy
Ensemble (Bagging)	Not Bullying	0.67	0.74	0.70	0.70
	Bullying	0.73	0.65	0.69	
Ensemble (Boosting)	Not Bullying	0.65	0.78	0.71	0.69
	Bullying	0.75	0.60	0.67	
Ensemble (Voting)	Not Bullying	0.69	0.73	0.71	<b>0.72</b>
	Bullying	0.74	0.70	0.72	

#### 4.4. Deep Learning for Multilingual Cyberbullying Detection

In this experiment, where we utilized a 70%:10%:20% training–validation–testing ratio, distinct performance trends emerged among various deep learning models. Based on the specified hyperparameters—Adam optimizer, a learning rate of 0.01, dropout set to

0.1, 5 epochs, and a batch size of 64—the assessment of model performance for detecting cyberbullying across Bangla and Chittagonian texts offered intriguing insights.

The Multilayer Perceptron (MLP) demonstrated moderate effectiveness, thus achieving a precision of 0.67 for nonbullying instances and 0.70 for bullying instances, with corresponding recall values of 0.68 and 0.69, respectively. However, its F1 scores and overall accuracy remained relatively modest at 0.68 and 0.69, respectively.

Moving to Recurrent Neural Network (RNN) architectures, the SimpleRNN, LSTM, and BiLSTM models exhibited notable enhancements in performance. Particularly, the LSTM model excelled with remarkable precision, recall, and F1 score values for both bullying and nonbullying instances, thus improving the highest accuracy among RNNs and showcasing its adeptness in capturing intricate linguistic patterns in cyberbullying detection tasks.

Similarly, the Gated Recurrent Unit (GRU) architecture delivered competitive results, with well-balanced precision, recall, and F1 score values for both bullying and nonbullying classes, thus mirroring its efficacy in classifying cyberbullying texts.

The Convolutional Neural Network (CNN) emerged as the most effective architecture, demonstrating high precision, recall, and F1 score values for both bullying and nonbullying instances. With precision scores of 0.79 and 0.83 for nonbullying and bullying instances, respectively, along with a recall of 0.82 for nonbullying and 0.80 for bullying instances, the CNN model showcased superior classification capabilities. Its F1 score of 0.80 for nonbullying and 0.81 for bullying instances further underscores its efficacy in cyberbullying detection tasks, as suggested in previous research [71]. Moreover, the CNN model achieved the highest overall accuracy of 0.811, thereby affirming its robustness in identifying cyberbullying content across Bangla and Chittagonian texts (refer to Table 4).

In conclusion, all DL models demonstrated potential in detecting cyberbullying; however, the LSTM, BiLSTM, GRU, and CNN architectures outperformed the others, with the CNN model being especially effective at capturing the nuances of cyberbullying in many languages.

**Table 4.** Deep learning model performance.

Model	Class	Precision	Recall	F1 Score	Accuracy
MLP	Not Bullying	0.67	0.68	0.68	0.69
	Bullying	0.70	0.69	0.70	
SimpleRNN	Not Bullying	0.74	0.76	0.75	0.76
	Bullying	0.77	0.76	0.77	
LSTM	Not Bullying	0.81	0.79	0.80	0.81
	Bullying	0.81	0.83	0.82	
BiLSTM	Not Bullying	0.79	0.77	0.78	0.79
	Bullying	0.79	0.81	0.80	
GRU	Not Bullying	0.79	0.80	0.79	0.80
	Bullying	0.81	0.80	0.81	
CNN	Not Bullying	0.79	0.82	0.80	<b>0.811</b>
	Bullying	0.83	0.80	0.81	

#### 4.5. Hybrid Deep Learning for Multilingual Cyberbullying Detection

In this experiment, we rigorously assessed our proposed models, thereby maintaining a consistent data splitting ratio of 70% for training, 10% for validation, and 20% for testing. Hybrid models, configured with an epoch setting of five and a batch size of 64, demonstrated compelling performance metrics. In the exploration of hybrid models for multilingual cyberbullying detection in Bangla and Chittagonian texts, we conducted a thorough analysis of various model architectures and evaluated their performance using precision, recall, F1 score, and accuracy metrics.

In the results, with an overall accuracy of 0.799, the BiLSTM+GRU model demonstrated balanced precision, recall, and F1 scores for both bullying and nonbullying classes.

This implies that the model (see Table 5) successfully captures minor linguistic patterns suggestive of cyberbullying behavior across a variety of textual data.

In contrast, the CNN+LSTM model demonstrated high precision for nonbullying instances but lower recall, thus resulting in a balanced F1 score of 0.78. However, its ability to identify bullying instances with a recall of 0.87 contributed to an overall accuracy of 0.801, thus indicating proficient detection capabilities.

The CNN+GRU model showcased balanced performance metrics, similar to the BiLSTM+GRU model, with an accuracy of 0.804. This suggests robust discrimination between cyberbullying and nonbullying content, thereby underscoring its potential for effective multilingual cyberbullying detection.

On the other hand, the CNN+BiLSTM model demonstrated higher precision for bullying instances but lower recall for nonbullying instances, thus resulting in an accuracy of 0.78. This indicates a need for further refinement to ensure balanced performance across both classes.

Lastly, the (CNN+LSTM)+BiLSTM hybrid model achieved balanced precision, recall, and F1 scores for both classes, culminating in an accuracy of 0.82. This highlights the model's potential for strong detection capabilities by demonstrating its ability to capture the subtleties of cyberbullying behavior in Bangla and Chittagonian texts.

**Table 5.** Hybrid model performance.

Model	Class	Precision	Recall	F1 Score	Accuracy
BiLSTM+GRU	Not Bullying	0.79	0.79	0.79	0.799
	Bullying	0.81	0.81	0.81	
CNN+LSTM	Not Bullying	0.84	0.72	0.78	0.801
	Bullying	0.77	0.87	0.82	
CNN+GRU	Not Bullying	0.79	0.80	0.80	0.804
	Bullying	0.81	0.81	0.81	
CNN+BiLSTM	Not Bullying	0.82	0.70	0.76	0.78
	Bullying	0.76	0.85	0.80	
<b>(CNN+LSTM)+BiLSTM</b>	Not Bullying	0.80	0.83	0.81	<b>0.82</b>
	Bullying	0.84	0.81	0.82	

#### 4.6. Transformer-Based Models for Multilingual Cyberbullying Detection

The results presented in this section showcase the performance of various transformer models in the multilingual cyberbullying detection task for Bangla and Chittagonian texts. Across the evaluated models, XLM-Roberta demonstrated notable precision and recall metrics for identifying instances of cyberbullying. With a precision of 0.78 and a recall of 0.91 for bullying instances, the XLM-Roberta model exhibited a strong capability to accurately classify instances of cyberbullying, thus achieving an overall accuracy of 0.841 (See Table 6 and Figure 9). This indicates that the model is a promising candidate for cyberbullying detection tasks, since it successfully detects instances of cyberbullying while maintaining a comparatively low False Positive rate. Figure 10 shows the confusion matrix of XLM-Roberta model.

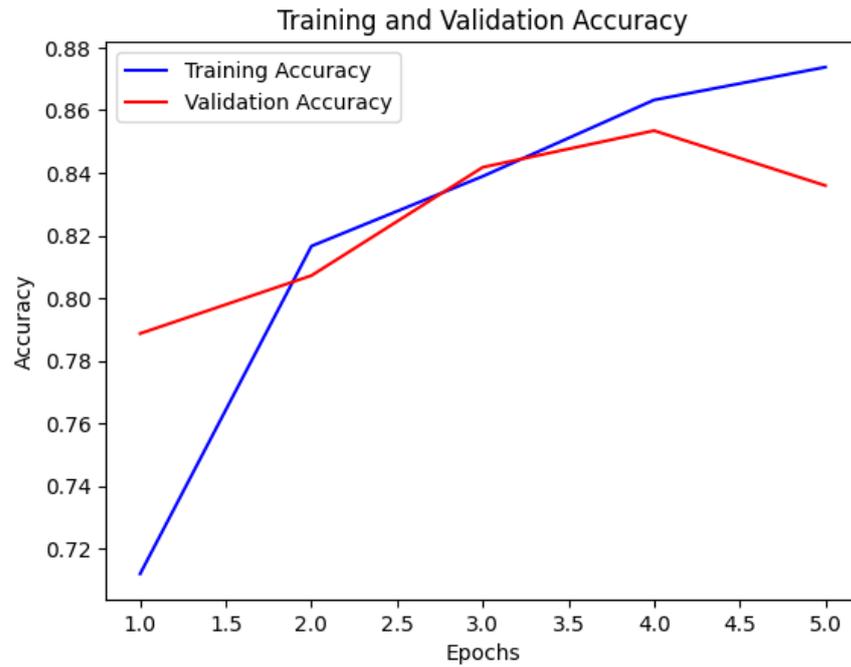


Figure 9. XLM-Roberta performance curve.

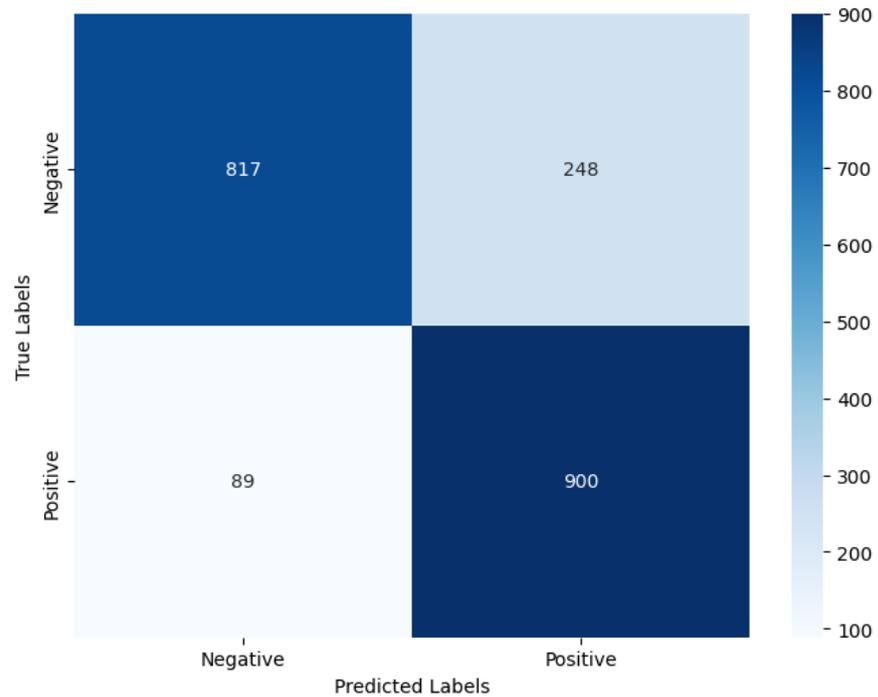


Figure 10. XLM-Roberta confusion matrix.

In contrast, the Bangla BERT and Multilingual BERT models exhibited consistent and balanced performance across both bullying and nonbullying classes, with the precision, recall, F1 score, and accuracy all reaching 0.82 ( See Figure 11). This consistency implies that these models maintained a stable performance across different evaluation metrics, thereby indicating robustness in their cyberbullying detection capabilities. Figure 12 shows the confusion matrix of the Bangla BERT model. Figure 13 shows the performance curve of Multilingual BERT based on the confusion matrix (see Figure 14). The BERT model also demonstrated balanced performance, with precision, recall, F1 score, and accuracy all at 0.82

for both bullying and nonbullying instances (See Figure 15) based on the confusion matrix (See Figure 16).

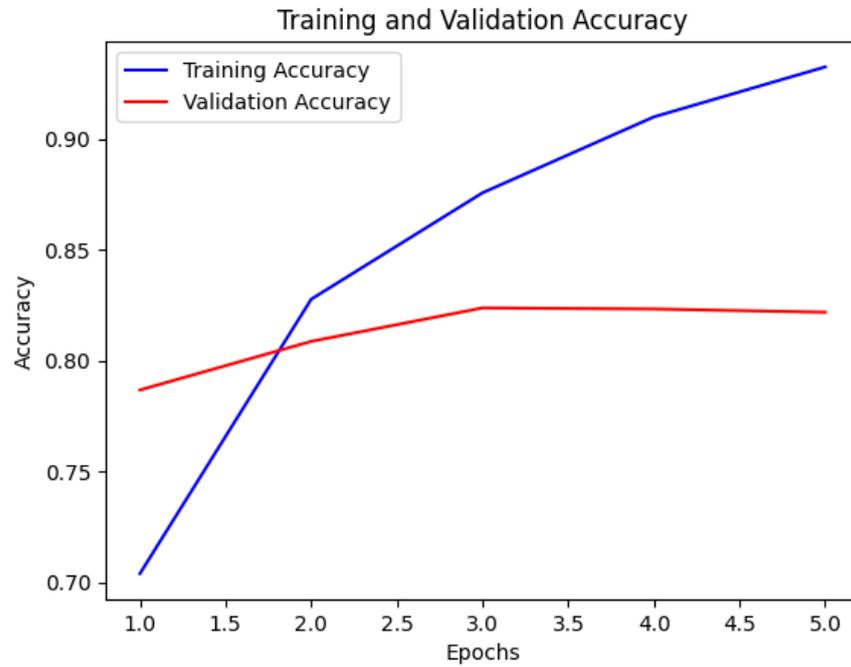


Figure 11. Bangla BERT performance curve.

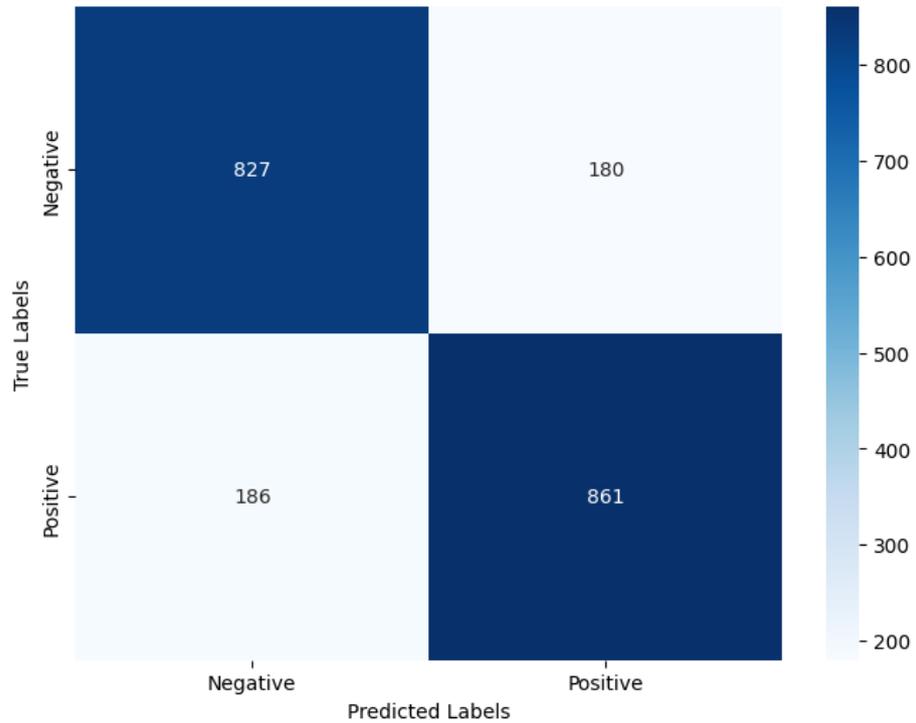


Figure 12. Bangla BERT confusion matrix.

However, the Bangla ELECTRA model showed a trade-off between precision and recall, particularly in identifying instances of cyberbullying. While achieving a high recall of 0.89 for bullying instances, the model’s precision was comparatively lower at 0.74, thus resulting in a lower F1 score of 0.81 and an accuracy of 0.785 (See Figure 17). This shows that although the model successfully captured a sizable percentage of cyberbullying incidents, it at the same time incorrectly categorized a sizable portion of nonbullying incidents,

which lowered the overall performance. Figure 18 shows the confusion matrix of Bangla ELECTRA model.

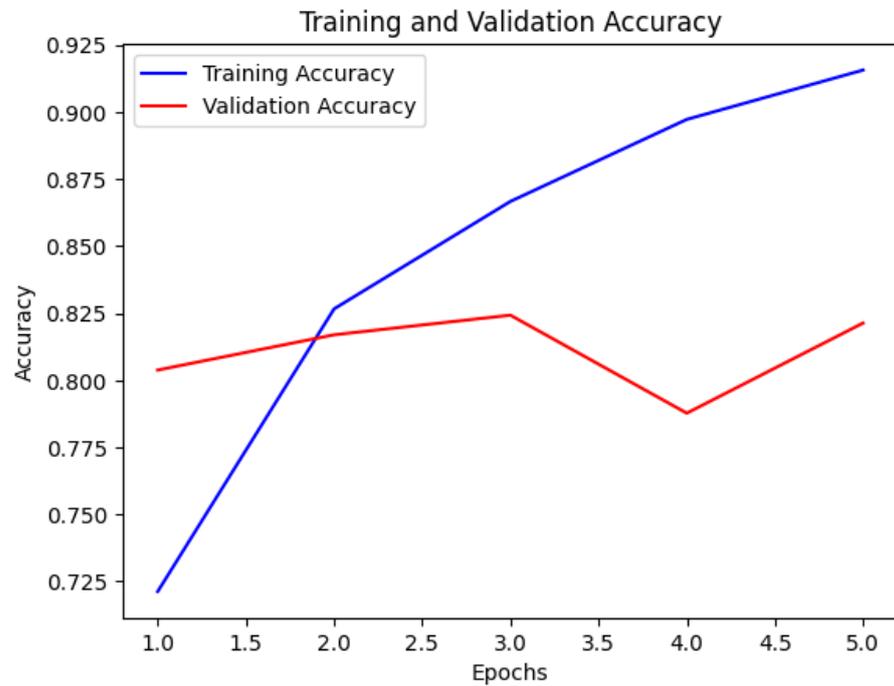


Figure 13. Multilingual BERT performance curve.

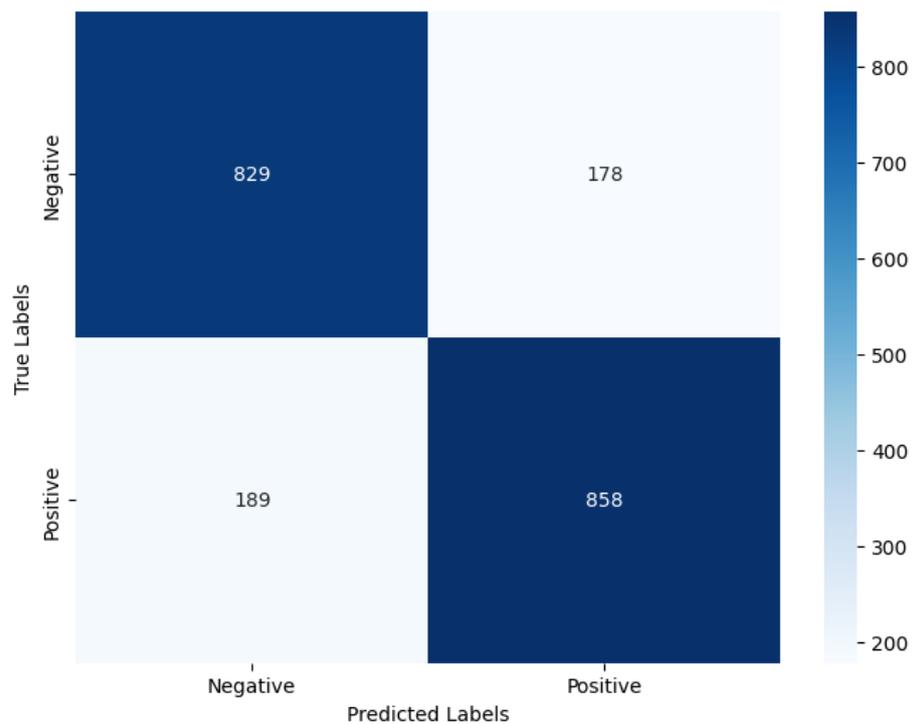


Figure 14. Multilingual BERT confusion matrix.

In general, the assessment highlights the effectiveness of transformer models in multilingual cyberbullying detection tasks in Bangla and Chittagonian texts, especially for XLM-Roberta, Bangla BERT, and Multilingual BERT. Among efficient cyberbullying detection techniques, these models exhibit strong performance metrics across a range of

evaluation criteria, thereby indicating their potential to improve online safety and promote healthier digital communities.

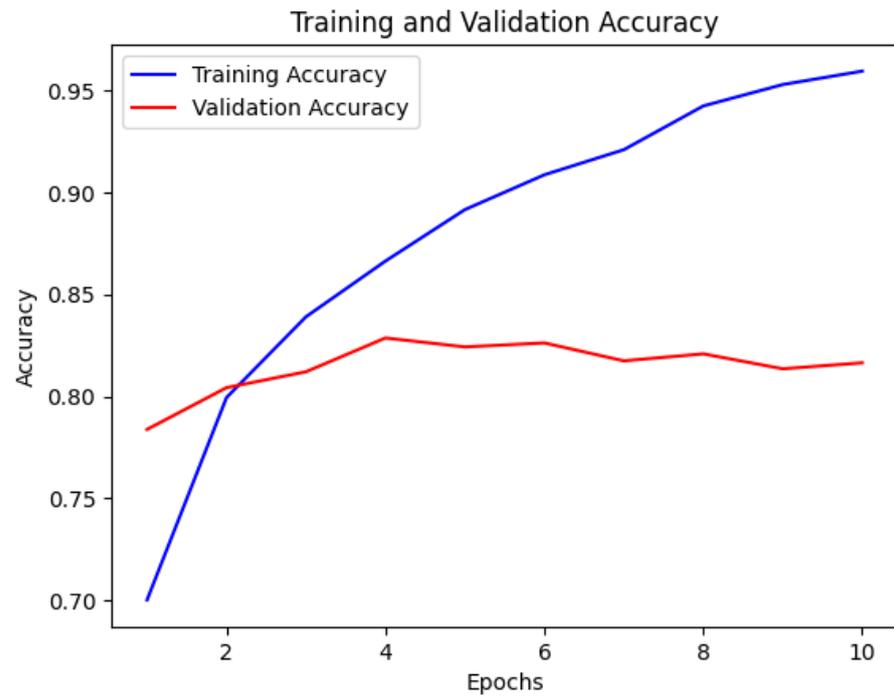


Figure 15. BERT performance curve.

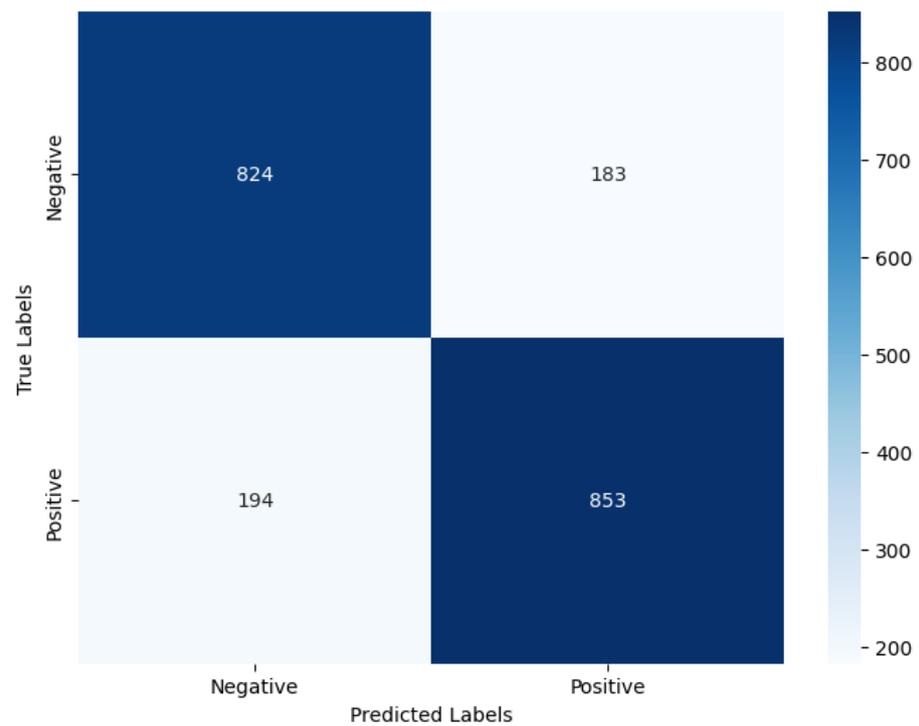


Figure 16. BERT confusion matrix.

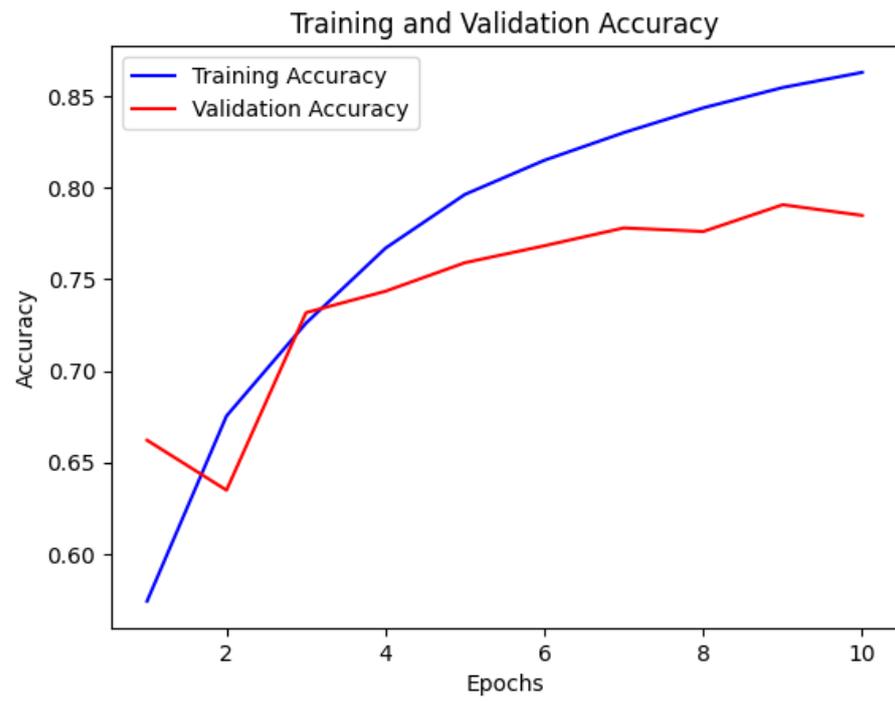


Figure 17. Bangla ELECTRA performance curve.

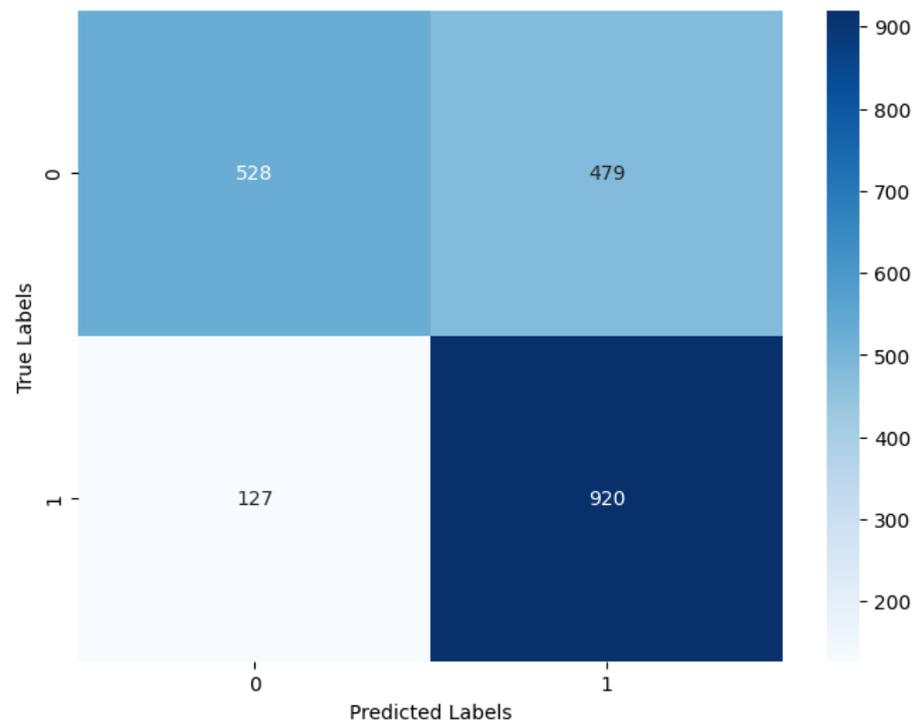


Figure 18. Bangla ELECTRA confusion matrix.

**Table 6.** Transformers model performance.

Model	Class	Precision	Recall	F1 Score	Accuracy
XLM-Roberta	Not Bullying	0.90	0.77	0.83	<b>0.841</b>
	Bullying	0.78	0.91	0.84	
Bangla BERT	Not Bullying	0.82	0.82	0.82	0.822
	Bullying	0.83	0.82	0.82	
Multilingual BERT	Not Bullying	0.81	0.82	0.82	0.821
	Bullying	0.83	0.82	0.82	
BERT	Not Bullying	0.81	0.82	0.81	0.820
	Bullying	0.82	0.81	0.82	
Bangla ELECTRA	Not Bullying	0.85	0.68	0.76	0.785
	Bullying	0.74	0.89	0.81	

#### 4.7. Comparative Analysis

The comparative analysis of machine learning and deep learning models for multilingual cyberbullying detection in Bangla and Chittagonian texts revealed distinctive performance trends across various algorithms and architectures. With balanced datasets of over 5000 samples each in Chittagonian and Bangla, covering cyberbullying and noncyberbullying instances, our study provides a robust evaluation framework.

In terms of machine learning algorithms, Support Vector Machine (SVM) emerged as the most effective classic ML model, thus achieving an accuracy of 0.711. Multinomial Naive Bayes (MNB) closely followed SVM with an accuracy of 0.71. Logistic Regression (LR) and Random Forest (RF) demonstrated moderate performance, while Decision Trees (DTs) and K-Nearest Neighbors (KNNs) performed slightly lower, with an accuracy of 0.63 as shown in Table 2.

The ensemble methods, including Bagging, Boosting, and Voting, showcased a commendable balance between precision, recall, F1 score, and accuracy. Bagging and Boosting exhibited similar trends, while Voting consistently outperformed both, thus achieving superior performance across all metrics as shown in Table 3.

Transitioning to deep learning architectures, the Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and Convolutional Neural Network (CNN) models emerged as top performers. The CNN model achieved the highest accuracy of 0.811, thus highlighting its proficiency in extracting valuable features from textual data shown in Table 4. Other models like BiLSTM+GRU, CNN+LSTM, CNN+GRU, CNN+BiLSTM, and (CNN+LSTM)+BiLSTM demonstrated balanced performance metrics, albeit with slight variations in precision, recall, and F1 score values, as shown in Table 5.

Some minor differences in the results between single-type networks and hybrid networks could come from two causes, namely, model architecture and hardware limitations. Regarding the first possible cause, namely the difference in architecture and how the networks are filtering out the features along the way, since CNNs and RNNs (including BiLSTMs), filter out features that are considered not important during their processing and use in classifications only those features that are considered important during the whole training run, the main difficulty in applying classic NNs has been in specifying the optimal architecture, such as the number of layers, types of layers, etc. Too many layers of a certain type could cause some useful features to be discarded along the way. Thus, a combined CNN+BiLSTM network could end up having lower results than the CNN-only network. Regarding the second possible cause, namely, hardware limitations, due to the limitations of the environment we were able to use for this research (Google Colaboratory free version), we needed to use different training parameters for single-type and hybrid networks (see Section 4.1), which could have also slightly influenced the results. However, since the overall differences in the results, especially the top-scoring networks, are not that large, it is not expected that the end result would be much different in a more robust environment (e.g., Google Colaboratory paid version).

Among Transformer-based models, XLM-Roberta demonstrated strong precision and recall metrics for cyberbullying detection, thus achieving an accuracy of 0.841, as shown in Table 6. The Bangla BERT, Multilingual BERT, and BERT models exhibited consistent and balanced performance across different evaluation metrics. However, Bangla ELECTRA showed a trade-off between precision and recall, thus indicating the need for further refinement.

As a whole, the comparative analysis shows how well different Transformer-based models, deep learning architectures, and machine learning algorithms identify cyberbullying samples in multilingual contexts. The research provides significant insights into the performance attributes of various models, thereby facilitating the identification of suitable approaches for cyberbullying detection tasks in Bangla and Chittagonian texts.

In fact, the datasets used in our study for Bangla and Chittagonian were balanced, with roughly equal numbers of texts falling into the categories of “bullying” and “not bullying”. In the Bangla dataset, 2545 instances were classified as bullying and 2555 instances were classified as nonbullying. Similarly, there were 2610 bullying instances and 2490 nonbullying instances in the Chittagonian dataset (See Figure 5). During training, the classifiers were exposed to an equal number of instances from each class, thus allowing them to learn effectively from both bullying and nonbullying examples.

We made sure that every class—bullying and nonbullying—had an equal representation during classifier training because of the datasets’ balanced nature. This strategy aids in avoiding problems with class imbalance, which may result in biased model performance.

#### 4.8. Comparison with Previous Studies

As far as we are aware, this study is the first to detect cyberbullying using multilingual models in Bangla and Chittagonian languages. One previous study by Mahmud et al. [54] also focused on Bangla and Chittagonian languages for cyberbullying detection. Below, we list the key distinctions and conclusions from the earlier research with reference to the present study:

1. Our creation of a dataset comprising over 5000 samples for both Bangla and Chittagonian languages, alongside the thorough validation, presents a unique contribution to linguistic research due to the absence of such datasets previously or prior investigations into the distinctions between the two languages [28,36,38–41].
2. Our study encompasses the Chittagonian and Bangla language domains, whereas previous research [81–84] was limited to the Bengali/Bangla language. Managing a multilingual model presented a number of challenges, including problems with data collection, annotation, processing, validation, and model construction. Despite these challenges, we overcame them to complete the study.
3. The study was implemented using five different sets of methods. Firstly, we evaluated the efficacy of traditional machine learning. We then investigated ensemble learning and deep learning methods. We also looked at hybrid deep learning techniques, and lastly, Transformer-based models. This all-encompassing strategy that incorporated different approaches represents the most extensive and exhaustive study so far when compared to all earlier research [15,16,18,37–41].
4. Upon building a variety of learning models and performing comparative analyses using various techniques, we observed that our models produced better results than those described in earlier studies [15,38,39].
5. Finally, we also paid much attention to the data annotation methodology (see Section 3.2) and ethical issues (see Section 4.9) regarding data collection, which previous studies did not account for [15,16,18,37–41].

#### 4.9. Ethical Considerations

In conducting this study, which revolved around the human population, particularly social media users and the unethical use of language online, an ethical approach to the research process was important—starting with data collection [85]. Data collection for this

study involved exclusively public user posts and comments. Facebook and other social media venues' open access policy permits the collection of data from public posts. Given that we gathered solely public comments/posts, they no longer retained their status as private, hence obviating the need for any special agreement for data collection and research implementation [86]. Furthermore, we adhered to stringent ethical standards during the collection of data from social media platforms, which included the following measures:

1. Data collection was limited to open SNS forums where users had previously consented to data collection for analytical purposes.
2. All identifying private information was removed from the collected data to ensure anonymity.
3. Data storage was conducted securely and exclusively for research purposes.
4. Materials that could potentially harm specific groups was avoided during the study.
5. Annotators were tasked with labeling the data conscientiously, considering gender and religious aspects of the data composition, with clear guidelines provided to mitigate bias during annotation.

## 5. Conclusions and Future Work

### 5.1. Conclusions

In this paper, we conducted a comprehensive investigation into multilingual cyberbullying detection focusing on Bangla and Chittagonian texts. Our study has yielded several noteworthy findings and implications for future research.

Firstly, through meticulous data collection and annotation processes, we curated a robust dataset comprising over 5000 samples from both Bangla and Chittagonian languages, thereby ensuring balanced representation. The high-quality ground truth data, validated using Krippendorff's alpha and Cohen's kappa scores, forms a solid foundation for cyberbullying detection research in these linguistic contexts.

Our analysis encompassed a wide range of methods, beginning with traditional machine learning and deep learning models. While traditional ML models, such as Support Vector Machine and ensemble methods showed promising results, DL models, particularly Convolutional Neural Networks, exhibited superior performance in the cyberbullying detection task.

Furthermore, we proposed and evaluated hybrid network-based models, thus demonstrating the effectiveness of combining different architectures for enhanced accuracy. Transformer-based models, including XLM-Roberta, Bangla BERT, and Multilingual BERT, also showed significant improvements in cyberbullying detection accuracy.

### 5.2. Contributions of this Study

The primary contributions of this study are as follows:

1. The dataset was meticulously compiled, thus containing manually gathered samples from both Bangla and Chittagonian languages. Each language contributed over 5000 samples, thus ensuring a balanced representation in the dataset.
2. Thorough manual labeling by human annotators guaranteed the production of quality ground truth data for both model training and evaluation. This quality was further verified through the rigorous assessment of Krippendorff's alpha and Cohen's kappa scores.
3. We conducted an extensive comparison of the performance of traditional machine learning (ML) models, with Support Vector Machine (SVM) emerging as the top performer, thus achieving an accuracy of 0.711.
4. We evaluated ensemble models including Bagging (0.70 accuracy), Boosting (0.69 accuracy), and Voting (0.72 accuracy), thus demonstrating promising results.
5. We explored Deep Learning (DL) models, particularly Convolutional Neural Network (CNN), which outperformed traditional ML approaches with accuracies ranging from 0.69 to 0.811.

6. We proposed a series of hybrid network-based models, such as BiLSTM+GRU, CNN+LSTM, CNN+BiLSTM, and CNN+GRU, thus achieving accuracies ranging from 0.78 to 0.804.
7. We investigated Transformer-based models, including XLM-Roberta, Bangla BERT, Multilingual BERT, BERT, and Bangla ELECTRA, which significantly enhanced accuracy levels.

### 5.3. Future Work

Future studies will concentrate on improving and enhancing current models; we intend to investigate various approaches to address the problem of data scarcity, such as data augmentation methods, transfer learning from similar tasks or languages, and active learning methods to iteratively obtain labeled data. Additionally, we will carry out thorough analyses to spot any biases in our datasets and create mitigation plans to enhance the generalization of our models in a variety of linguistic and cultural contexts.

Enhancing the efficacy of cyberbullying detection systems can also be accomplished through the creation of specialized tools and resources, such as NER tools for Chittagonian. Additionally, investigating explainable AI (XAI) techniques will enhance model interpretability, thus fostering safer online environments.

**Author Contributions:** Conceptualization, T.M. and M.P.; methodology, T.M.; validation, T.M., M.P., and F.M.; formal analysis, T.M. and M.P.; investigation, T.M. and M.P.; resources, T.M. and M.P.; data curation, T.M. and M.P.; writing—original draft preparation, T.M.; writing—review and editing, T.M. and M.P.; visualization, T.M.; supervision, M.P. and F.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** All necessary informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data used to support the findings of this study are available upon reasonable request to the corresponding author.

**Acknowledgments:** This study was carried out at the Text Information Processing Laboratory (TIP Lab) at the Kitami Institute of Technology located in Kitami, Hokkaido, Japan. The authors express their gratitude to the lead supervisor for providing guidance, continuous evaluation, and oversight throughout the research process. Special acknowledgment is given to the dedicated three annotators whose active involvement significantly contributed to annotating and refining the dataset. Additionally, we utilized language support tools for refining grammar, which greatly enhanced the linguistic quality of the manuscript.

**Conflicts of Interest:** The authors declare no conflicts of interest.

### Abbreviations

The following abbreviations are used in this manuscript:

NLP	Natural Language Processing
ML	Machine Learning
DL	Deep Learning
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
BTRC	Bangladesh Telecommunication Regulatory Commission
GRU	Gated Recurrent Unit
TF-IDF	Term Frequency–Inverse Document Frequency
MLP	Multilayer Perceptron
LR	Logistic Regression

GBM	Gradient Boosting Machine
SVM	Support Vector Machine
KNN	K-Nearest Neighbors
DT	Decision Tree
RF	Random Forest
MNB	Multinomial Naive Bayes
LSTM	Long Short-Term Memory network
BiLSTM	Bidirectional Long Short-Term Memory network
BERT	Bidirectional Encoder Representations from Transformers
mBERT	Multilingual BERT
ELECTRA	Pretraining Text Encoders as Discriminators Rather Than Generators
XAI	Explainable AI

## References

1. Bangladesh Telecommunication Regulatory Commission. Available online: <http://www.btrc.gov.bd/site/page/347df7fe-409f-451e-a415-65b109a207f5/-> (accessed on 15 January 2023).
2. United Nations Development Programme. Available online: <https://www.undp.org/bangladesh/blog/digital-bangladesh-innovative-bangladesh-road-2041> (accessed on 20 January 2023).
3. Chittagong City in Bangladesh. Available online: <https://en.wikipedia.org/wiki/Chittagong> (accessed on 1 April 2023).
4. StatCounter Global Stats. Available online: <https://gs.statcounter.com/social-media-stats/all/bangladesh/#monthly-202203-202303> (accessed on 28 April 2023).
5. Chittagonian Language. Available online: [https://en.wikipedia.org/wiki/Chittagonian\\_language](https://en.wikipedia.org/wiki/Chittagonian_language) (accessed on 11 February 2023).
6. Bengali Language. Available online: <https://en.wikipedia.org/wiki/Bengalilanguage> (accessed on 11 February 2023).
7. Mahmud, T.; Ptaszynski, M.; Eronen, J.; Masui, F. Cyberbullying detection for low-resource languages and dialects: Review of the state of the art. *Inf. Process. Manag.* **2023**, *60*, 103454. [CrossRef]
8. Mahmud, T.; Ptaszynski, M.; Masui, F. Automatic Vulgar Word Extraction Method with Application to Vulgar Remark Detection in Chittagonian Dialect of Bangla. *Appl. Sci.* **2023**, *13*, 11875. [CrossRef]
9. Facebook. Available online: <https://www.facebook.com/> (accessed on 28 January 2023).
10. imo. Available online: <https://imo.im> (accessed on 28 January 2023).
11. WhatsApp. Available online: <https://www.whatsapp.com> (accessed on 28 January 2023).
12. Addiction Center. Available online: <https://www.addictioncenter.com/drugs/social-media-addiction/> (accessed on 28 January 2023).
13. Prothom Alo. Available online: <https://en.prothomalo.com/bangladesh/Youth-spend-80-mins-a-day-in-Internet-adda> (accessed on 28 January 2023).
14. United Nations. Available online: <https://www.un.org/en/chronicle/article/cyberbullying-and-its-implications-human-rights> (accessed on 28 January 2023).
15. Pawar, R.; Raje, R.R. Multilingual cyberbullying detection system. In Proceedings of the 2019 IEEE international conference on electro information technology (EIT), Brookings, SD, USA, 20–22 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 040–044.
16. Haidar, B.; Chamoun, M.; Serhrouchni, A. Multilingual cyberbullying detection system: Detecting cyberbullying in Arabic content. In Proceedings of the 2017 1st cyber security in networking conference (CSNet), Rio de Janeiro, Brazil, 18–20 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–8.
17. Okoloegbo, C.A.; Eze, U.F.; Chukwudebe, G.A.; Nwokonkwo, O.C. Multilingual Cyberbullying Detector (CD) Application for Nigerian Pidgin and Igbo Language Corpus. In Proceedings of the 2022 5th Information Technology for Education and Development (ITED), Abuja, Nigeria, 1–3 November 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–6.
18. Mahajan, E.; Mahajan, H.; Kumar, S. EnsMulHateCyb: Multilingual hate speech and cyberbully detection in online social media. *Expert Syst. Appl.* **2023**, *236*, 121228. [CrossRef]
19. Mahmud, T.; Barua, A.; Begum, M.; Chakma, E.; Das, S.; Sharmen, N. An Improved Framework for Reliable Cardiovascular Disease Prediction Using Hybrid Ensemble Learning. In Proceedings of the 2023 International Conference on Electrical, Computer and Communication Engineering (ECCE), Chittagong, Bangladesh, 23–25 February 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.
20. Mahmud, T.; Barua, A.; Islam, D.; Hossain, M.S.; Chakma, R.; Barua, K.; Monju, M.; Andersson, K. Ensemble Deep Learning Approach for ECG-Based Cardiac Disease Detection: Signal and Image Analysis. In Proceedings of the 2023 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), Dhaka, Bangladesh, 21–23 September 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 70–74.
21. Nuha, N.S.; Mahmud, T.; Rezaoana, N.; Hossain, M.S.; Andersson, K. An Approach of Analyzing Classroom Student Engagement in Multimodal Environment by Using Deep Learning. In Proceedings of the 2023 IEEE 9th International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), Thiruvananthapuram, India, 25–26 November 2023; pp. 286–291. [CrossRef]
22. Das, S.; Mahmud, T.; Islam, D.; Begum, M.; Barua, A.; Tarek Aziz, M.; Nur Showan, E.; Dey, L.; Chakma, E. Deep Transfer Learning-Based Foot No-Ball Detection in Live Cricket Match. *Comput. Intell. Neurosci.* **2023**, *2023*, 2398121. [CrossRef] [PubMed]

23. Barua, K.; Mahmud, T.; Barua, A.; Sharmen, N.; Basnin, N.; Islam, D.; Hossain, M.S.; Andersson, K.; Hossain, S. Explainable AI-Based Humerus Fracture Detection and Classification from X-Ray Images. In Proceedings of the 2023 26th International Conference on Computer and Information Technology (ICIT), Cox's Bazar, Bangladesh, 13–15 December 2023; pp. 1–6. [CrossRef]
24. Dey, P.; Mahmud, T.; Nahar, S.R.; Hossain, M.S.; Andersson, K. Plant Disease Detection in Precision Agriculture: Deep Learning Approaches. In Proceedings of the 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 4–6 January 2024; pp. 661–667. [CrossRef]
25. Mahmud, T.; Barua, K.; Barua, A.; Das, S.; Basnin, N.; Hossain, M.S.; Andersson, K.; Kaiser, M.S.; Sharmen, N. Exploring Deep Transfer Learning Ensemble for Improved Diagnosis and Classification of Alzheimer's Disease. In Proceedings of the International Conference on Brain Informatics, Hoboken, NJ, USA, 1–3 August 2023; Springer: Cham, Switzerland, 2023; pp. 109–120.
26. Chowdhury, N.A.; Mahmud, T.; Barua, A.; Basnin, N.; Barua, K.; Iqbal, A.; Hossain, M.S.; Andersson, K.; Kaiser, M.S.; Hossain, M.S.; et al. A Novel Approach to Detect Stroke from 2D Images Using Deep Learning. In Proceedings of the 2nd International Conference on Big Data, IoT and Machine Learning, Dhaka, Bangladesh, 6–8 September 2023; pp. 239–253.
27. Mahmud, T.; Barua, K.; Habiba, S.U.; Sharmen, N.; Hossain, M.S.; Andersson, K. An Explainable AI Paradigm for Alzheimer's Diagnosis Using Deep Transfer Learning. *Diagnostics* **2024**, *14*, 345. [CrossRef]
28. Malte, A.; Ratadiya, P. Multilingual cyber abuse detection using advanced transformer architecture. In Proceedings of the TENCON 2019–2019 IEEE Region 10 Conference (TENCON), Kochi, India, 17–20 October 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 784–789.
29. Krippendorff, K. Measuring the reliability of qualitative text analysis data. *Qual. Quant.* **2004**, *38*, 787–800. [CrossRef]
30. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [CrossRef]
31. Ptaszynski, M.; Dybala, P.; Matsuba, T.; Masui, F.; Rzepka, R.; Araki, K. Machine learning and affect analysis against cyber-bullying. In Proceedings of the 36th AISB, Leicester, UK, 29 March–1 April 2010; pp. 7–16.
32. Ptaszynski, M.; Dybala, P.; Matsuba, T.; Masui, F.; Rzepka, R.; Araki, K.; Momouchi, Y. In the service of online order: Tackling cyber-bullying with machine learning and affect analysis. *Int. J. Comput. Linguist. Res.* **2010**, *1*, 135–154.
33. Ptaszynski, M.; Pieciukiewicz, A.; Dybala, P. Results of the PolEval 2019 Shared Task 6: First Dataset and Open Shared Task for Automatic Cyberbullying Detection in Polish Twitter. In Proceedings of the PolEval 2019 Workshop, Warsaw, Poland, 31 May 2019; p. 89.
34. Eronen, J.; Ptaszynski, M.; Masui, F.; Smywiński-Pohl, A.; Leliwa, G.; Wroczynski, M. Improving classifier training efficiency for automatic cyberbullying detection with feature density. *Inf. Process. Manag.* **2021**, *58*, 102616. [CrossRef]
35. Eronen, J.; Ptaszynski, M.; Masui, F.; Arata, M.; Leliwa, G.; Wroczynski, M. Transfer language selection for zero-shot cross-lingual abusive language detection. *Inf. Process. Manag.* **2022**, *59*, 102981. [CrossRef]
36. Ranasinghe, T.; Zampieri, M. Multilingual offensive language identification for low-resource languages. *Trans. Asian Low-Resour. Lang. Inf. Process.* **2021**, *21*, 1–13. [CrossRef]
37. Corazza, M.; Menini, S.; Cabrio, E.; Tonelli, S.; Villata, S. A multilingual evaluation for online hate speech detection. *ACM Trans. Internet Technol. (TOIT)* **2020**, *20*, 1–22. [CrossRef]
38. Si, S.; Datta, A.; Banerjee, S.; Naskar, S.K. Aggression detection on multilingual social media text. In Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 6–8 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–5.
39. Roy, S.G.; Narayan, U.; Raha, T.; Abid, Z.; Varma, V. Leveraging multilingual transformers for hate speech detection. *arXiv* **2021**, arXiv:2101.03207.
40. El-Alami, F.z.; El Alaoui, S.O.; Nahnahi, N.E. A multilingual offensive language detection method based on transfer learning from transformer fine-tuning model. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 6048–6056. [CrossRef]
41. Das, A.; Nandy, S.; Saha, R.; Das, S.; Saha, D. Analysis and Detection of Multilingual Hate Speech Using Transformer Based Deep Learning. *arXiv* **2024**, arXiv:2401.11021.
42. Chatgaiya Express. Available online: <https://web.facebook.com/groups/1535657099839469> (accessed on 27 April 2022).
43. Chatgaiya Tourist Gang. Available online: <https://web.facebook.com/groups/999621230477551> (accessed on 7 May 2022).
44. Hardeniya, N.; Perkins, J.; Chopra, D.; Joshi, N.; Mathur, I. *Natural Language Processing: Python and NLTK*; Packt Publishing Ltd.: Birmingham, UK, 2016.
45. Millstein, F. *Natural Language Processing with Python: Natural Language Processing Using NLTK*; CreateSpace Independent Publishing Platform: Scotts Valley, CA, USA, 2020.
46. Ptaszynski, M.; Maciejewski, J.; Dybala, P.; Rzepka, R.; Araki, K. CAO: A fully automatic emoticon analysis system based on theory of kinesics. *IEEE Trans. Affect. Comput.* **2010**, *1*, 46–59. [CrossRef]
47. Li, D.; Rzepka, R.; Ptaszynski, M.; Araki, K. HEMOS: A novel deep learning-based fine-grained humor detecting method for sentiment analysis of social media. *Inf. Process. Manag.* **2020**, *57*, 102290. [CrossRef]
48. Haque, M.Z.; Zaman, S.; Saurav, J.R.; Haque, S.; Islam, M.S.; Amin, M.R. B-NER: A Novel Bangla Named Entity Recognition Dataset with Largest Entities and Its Baseline Evaluation. *IEEE Access* **2023**, *11*, 45194–45205. [CrossRef]
49. Aizawa, A. An information-theoretic perspective of tf-idf measures. *Inf. Process. Manag.* **2003**, *39*, 45–65. [CrossRef]

50. Chakraborty, M.; Huda, M.N. Bangla document categorisation using multilayer dense neural network with tf-idf. In Proceedings of the 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 3–5 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–4.
51. Mahmud, T.; Das, S.; Ptaszynski, M.; Hossain, M.S.; Andersson, K.; Barua, K. Reason Based Machine Learning Approach to Detect Bangla Abusive Social Media Comments. In *Intelligent Computing & Optimization, Proceedings of the 5th International Conference on Intelligent Computing and Optimization 2022 (ICO2022)*, Hua Hin, Thailand, 27–28 October 2022; Springer: Cham, Switzerland, 2022; pp. 489–498.
52. Rahman, R. Robust and consistent estimation of word embedding for bangla language by fine-tuning word2vec model. In Proceedings of the 2020 23rd International Conference on Computer and Information Technology (ICCIT), DHAKA, Bangladesh, 19–21 December 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
53. Mahmud, T.; Ptaszynski, M.; Masui, F. Vulgar Remarks Detection in Chittagonian Dialect of Bangla. *arXiv* **2023**, arXiv:2308.15448.
54. Mahmud, T.; Ptaszynski, M.; Masui, F. Deep Learning Hybrid Models for Multilingual Cyberbullying Detection: Insights from Bangla and Chittagonian Languages. In Proceedings of the 2023 26th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, 13–15 December 2023; pp. 1–6. [[CrossRef](#)]
55. Mahmud, T.; Hasan, I.; Aziz, M.T.; Rahman, T.; Hossain, M.S.; Andersson, K. Enhanced Fake News Detection through the Fusion of Deep Learning and Repeat Vector Representations. In Proceedings of the 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 4–6 January 2024; pp. 654–660. [[CrossRef](#)]
56. Schneider, K.M. A comparison of event models for Naive Bayes anti-spam e-mail filtering. In Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics, Budapest, Hungary, 12–17 April 2003.
57. Kadam, S.; Gala, A.; Gehlot, P.; Kurup, A.; Ghag, K. Word embedding based multinomial naive bayes algorithm for spam filtering. In Proceedings of the 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 16–18 August 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–5.
58. Szarvas, G.; Farkas, R.; Kocsor, A. A multilingual named entity recognition system using boosting and c4. 5 decision tree learning algorithms. In Proceedings of the Discovery Science: 9th International Conference, DS 2006, Barcelona, Spain, 7–10 October 2006; Proceedings 9; Springer: Cham, Switzerland, 2006; pp. 267–278.
59. Stap, D.; Monz, C. Multilingual k-Nearest-Neighbor Machine Translation. *arXiv* **2023**, arXiv:2310.14644.
60. Wadud, M.A.H.; Mridha, M.; Shin, J.; Nur, K.; Saha, A.K. Deep-BERT: Transfer Learning for Classifying Multilingual Offensive Texts on Social Media. *Comput. Syst. Sci. Eng.* **2023**, *44*, 1775–1791. [[CrossRef](#)]
61. Roy, P.K. Deep Ensemble Network for Sentiment Analysis in Bi-lingual Low-resource Languages. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2024**, *23*, 1–16. [[CrossRef](#)]
62. Sharma, P.; Parwekar, P. Comparing Ensemble Techniques for Bilingual Multiclass Classification of Online Reviews. In Proceedings of the International Conference On Emerging Trends In Expert Applications & Security, Jaipur, India, 17–18 February 2018; Springer: Singapore, 2023; pp. 275–284.
63. Saha, S.; Ekbal, A. Combining multiple classifiers using vote based classifier ensemble technique for named entity recognition. *Data Knowl. Eng.* **2013**, *85*, 15–39. [[CrossRef](#)]
64. Thomas, S.; Ganapathy, S.; Hermansky, H. Multilingual MLP features for low-resource LVCSR systems. In Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 25–30 March 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 4269–4272.
65. Can, E.F.; Ezen-Can, A.; Can, F. Multilingual sentiment analysis: An RNN-based framework for limited data. *arXiv* **2018**, arXiv:1806.04511.
66. Samih, Y.; Maharjan, S.; Attia, M.; Kallmeyer, L.; Solorio, T. Multilingual code-switching identification via lstm recurrent neural networks. In Proceedings of the Second Workshop on Computational Approaches to Code Switching, Austin, TX, USA, 1 November 2016; pp. 50–59.
67. Arslan, R.S.; Barışçı, N. Development of output correction methodology for long short term memory-based speech recognition. *Sustainability* **2019**, *11*, 4250. [[CrossRef](#)]
68. Nguyen, H.T.; Le Nguyen, M. Multilingual opinion mining on YouTube—A convolutional N-gram BiLSTM word embedding. *Inf. Process. Manag.* **2018**, *54*, 451–462. [[CrossRef](#)]
69. Zhang, M.; Geng, G.; Chen, J. Semi-supervised bidirectional long short-term memory and conditional random fields model for named-entity recognition using embeddings from language models representations. *Entropy* **2020**, *22*, 252. [[CrossRef](#)]
70. Mohammad, A.S.; Hammad, M.M.; Sa'ad, A.; Saja, A.T.; Cambria, E. Gated recurrent unit with multilingual universal sentence encoder for Arabic aspect-based sentiment analysis. *Knowl.-Based Syst.* **2023**, *261*, 107540.
71. Ptaszynski, M.; Eronen, J.K.K.; Masui, F. Learning Deep on Cyberbullying is Always Better Than Brute Force. In Proceedings of the LaCATODA@IJCAI, Melbourne, Australia, 21 August 2017; pp. 3–10.
72. Yadav, A.; Vishwakarma, D.K. A multilingual framework of CNN and bi-LSTM for emotion classification. In Proceedings of the 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 1–3 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
73. Artene, C.G.; Tibeică, M.N.; Leon, F. Using BERT for multi-label multi-language web page classification. In Proceedings of the 2021 IEEE 17th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 28–30 October 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 307–312.

74. Khan, F.; Mustafa, R.; Tasnim, F.; Mahmud, T.; Hossain, M.S.; Andersson, K. Exploring BERT and ELMo for Bangla Spam SMS Dataset Creation and Detection. In Proceedings of the 2023 26th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, 13–15 December 2023; pp. 1–6. [[CrossRef](#)]
75. Kowsher, M.; Sami, A.A.; Prottasha, N.J.; Arefin, M.S.; Dhar, P.K.; Koshiba, T. Bangla-BERT: Transformer-based efficient model for transfer learning and language understanding. *IEEE Access* **2022**, *10*, 91855–91870. [[CrossRef](#)]
76. Rahman, M.M.; Pramanik, M.A.; Sadik, R.; Roy, M.; Chakraborty, P. Bangla documents classification using transformer based deep learning models. In Proceedings of the 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 19–20 December 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–5.
77. Wang, Z.; Mayhew, S.; Roth, D. Extending multilingual BERT to low-resource languages. *arXiv* **2020**, arXiv:2004.13640.
78. Ou, X.; Li, H. YNU@ Dravidian-CodeMix-FIRE2020: XLM-RoBERTa for Multi-language Sentiment Analysis. In Proceedings of the FIRE (Working Notes), Hyderabad, India, 16–20 December 2020; pp. 560–565.
79. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv* **2019**, arXiv:1907.11692.
80. Vujovic, Z. Classification model evaluation metrics. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*, 599–606. [[CrossRef](#)]
81. Islam, T.; Ahmed, N.; Latif, S. An evolutionary approach to comparative analysis of detecting Bangla abusive text. *Bull. Electr. Eng. Inform.* **2021**, *10*, 2163–2169. [[CrossRef](#)]
82. Karim, M.R.; Dey, S.K.; Islam, T.; Sarker, S.; Menon, M.H.; Hossain, K.; Hossain, M.A.; Decker, S. Deephateexplainer: Explainable hate speech detection in under-resourced bengali language. In Proceedings of the 2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA), Porto, Portugal, 6–9 October 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–10.
83. Ishmam, A.M.; Sharmin, S. Hateful speech detection in public facebook pages for the bengali language. In Proceedings of the 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 555–560.
84. Sazzed, S. Abusive content detection in transliterated Bengali-English social media corpus. In Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching, Online, 11 June 2021; pp. 125–130.
85. Gray, D.E. *Doing Research in the Real World*; Sage: Newcastle upon Tyne, UK, 2021; pp. 1–100.
86. Mahoney, J.; Le Louvier, K.; Lawson, S.; Bertel, D.; Ambrosetti, E. Ethical considerations in social media analytics in the context of migration: Lessons learned from a Horizon 2020 project. *Res. Ethics* **2022**, *18*, 226–240. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.