Multi-Source Bangla Violence Text Dataset and Transformer-Based Stacking Ensemble for Social Media Content Moderation

Anonymous ACL submission

Abstract

The exponential rise in user-generated content on social media platforms, such as Facebook, YouTube, and TikTok, has led to an alarming increase in the spread of violence-inciting language, especially in low-resource languages like Bangla. This issue has amplified the need for effective automated systems capable of detecting and filtering harmful content to ensure safer digital environments. In this study, we propose the BX Stacking Ensemble Model, a novel approach that combines the strengths of two powerful transformer-based models, BanglaBERT and XLM-RoBERTa, to improve the detection of violence-related text in Bangla. The model is trained on a newly compiled, diverse dataset of 11,933 samples, which includes both the Vio-Lens dataset and additional instances collected from social media platforms like YouTube, Facebook, and TikTok. The dataset is carefully annotated into three categories: Non-Violence, Passive Violence, and Active Violence. We compare the performance of the BX Stacking Ensemble Model with traditional machine learning models and other transformer-based models, demonstrating that the ensemble approach significantly outperforms baseline models, achieving a Macro F1 score of 0.85. The results highlight the effectiveness of combining both language-specific and multilingual transformers, enabling the detection of nuanced violence-inciting content. This research contributes to the development of more robust and scalable solutions for content moderation, particularly in resource-constrained languages like Bangla. Moreover, it demonstrates the potential of ensemble learning techniques in addressing the challenges of complex text classification tasks in real-world applications.

1 Introduction

In recent years, the exponential growth of social media platforms such as YouTube, TikTok, Facebook, Instagram, and Twitter has transformed the digital communication landscape. Millions of users now express their opinions, emotions, and experiences freely through posts, comments, and videos. However, this surge in user-generated content has also created an alarming rise in violent and hateful text online. This type of content not only promotes harassment and psychological harm but also fuels social polarization and extremism. As a result, automated violence text detection has become an essential research area within Natural Language Processing (NLP), aiming to ensure a safer and more inclusive online ecosystem. In South Asia, particularly in Bangladesh and West Bengal, social media has been weaponized to incite communal violence, as evidenced by incidents like the 2021 Cumilla Durga Puja violence triggered by a Facebook post (Saha et al., 2023b).

Despite progress in content moderation for high-resource languages like English, Bengali remains underrepresented in computational linguistics due to challenges such as rich morphology, code-mixing, slang, and cultural nuances. The rise of violent content on online platforms has sparked interest in automatic violence detection for low-resource languages like Bengali. The Vio-Lens dataset (Saha et al., 2023b) has been pivotal in advancing research, enabling diverse modeling strategies. Recent studies have explored both traditional and neural methods, with fine-tuned BanglaBERT revealing misclassification issues from inconsistent labeling (Khondoker et al., 2025), and ensemble methods improving robustness (Shibu et al., 2023). While traditional machine learning approaches have been compared (Alamgir and Haque, 2023) and transformer-based models demonstrated effectiveness (Dey et al., 2023), challenges remain due to dataset limitations, linguistic diversity, and informal social media text, highlighting the need for more scalable violence detection frameworks.

Our proposed ensemble-based model addresses the detection of violence-inciting text in Bengali so-

Figure 1: Word Cloud of the Dataset

cial media by leveraging a newly compiled dataset of 11,933 samples, combining the Vio-Lens corpus with 5,909 additional instances from YouTube, Facebook, and TikTok. The dataset is annotated into Non-Violence, Passive Violence, and Active Violence categories, enhancing its diversity and linguistic coverage. As illustrated by the word cloud in Figure 1, the dataset captures a broad range of terms, reflecting the varied nature of online aggression. The ensemble methodology, combining BERT models with machine learning and deep learning techniques, significantly improves classification metrics compared to single-model baselines. This model's scalability and adaptability make it suitable for real-world content moderation. In summary, this study provides a robust solution for violence detection in Bengali, balancing contextual understanding and model diversity. It includes both passive and active violence, offering a comprehensive analysis of online aggression. The research aims to contribute to safer digital communication in Bengali-speaking communities.

2 Literature Review

2.1 Traditional Machine Learning-Based Approaches

The detection of violence-inciting language in Bangla has been approached using various machine learning models, with early efforts focused on traditional methods. An ensemble-based classification system incorporating BanglaBERT, SagorBERT, and mBERT was proposed, effectively addressing challenges related to class imbalance and limited computational resources (Shibu et al., 2023). In a comparative evaluation of Bangla text classification approaches, both traditional machine learning and transformer-based models were analyzed, with particular emphasis placed on the impact of noisy data such as spelling variations and informal language, on the detection of violence-inciting content (Alamgir and Haque, 2023).

2.2 Transformer-Based Models

The advent of transformer-based models has significantly impacted the detection of violence-inciting language in Bangla. The fine-tuning of BanglaBERT on an expanded dataset was investigated, revealing difficulties in detecting subtle violence cues due to inconsistencies in annotation (Khondoker et al., 2025). Multilingual transformer models, including mBERT and XLM-RoBERTa, were evaluated, with domain-specific fine-tuning emphasized as essential despite the substantial computational costs associated with large-scale architectures (Dey et al., 2023). A two-stage classification pipeline leveraging XLM-R and MuRIL was employed to mitigate overfitting arising from dataset imbalance (Raihan et al., 2023). Additionally, multiple transformer-based models were benchmarked, with further hyperparameter optimization recommended to improve overall performance (Saha and Nanda, 2023). While strong performance was observed for multilingual models such as XLM-RoBERTa, limitations in capturing Bangla-specific linguistic characteristics were also reported (Mukherjee et al., 2023).

2.3 Hybrid or Combined Techniques

Efforts to improve violence detection in Bangla have also focused on combining different techniques and approaches. The BASED dataset was introduced to expand the linguistic diversity of Bangla violence detection, although challenges related to oversimplified class definitions were noted (Tasnim et al., 2024). Detection of political cyberbullying in Bangla was studied through the classification of large-scale Facebook comments (Ahmed et al., 2023). A Hierarchical-BERT model was proposed to address issues of class imbalance and linguistic complexity (Das et al., 2023). Performance improvements for BanglaBERT were achieved through self-training and data augmentation, mitigating the effects of data scarcity and imbalance in violence detection (Riyad et al., 2023). Furthermore, the VITD task at BLP-2023 was introduced, providing new benchmarks for violence-inciting text classification in Bangla (Saha et al., 2023a). An ensemble-based approach combining BanglaBERT, MuRIL, and XLM-RoBERTa was employed to explore the potential of ensemble methods in enhancing detection accuracy (Page et al., 2023).

3 Dataset Development

3.1 Overview of the Dataset

In this study, we constructed a larger and more diverse Bangla text dataset comprising 11,933 samples to train and evaluate models for detecting violence-inciting content on social media. The dataset consists of real user comments and posts in Bangla and is annotated into three categories: *Non-Violence* (peaceful or harmless statements), *Passive Violence* (indirect or subtle promotion of aggression, discrimination, or harmful ideologies, including sarcasm, ridicule, or justification of violence), and *Active Violence* (direct threats, abuse, or physical aggression). These categories collectively capture the full spectrum of harmful online language. Each sample was independently labeled by three annotators following predefined guidelines, and final labels were assigned through majority voting after reviewing ambiguous cases. The reliability of the annotation process was assessed using Fleiss' Kappa, achieving an inter-annotator agreement score of 0.89, which indicates strong labeling consistency (Fleiss, 1971).

Dataset Source	NV	PV	AV
Vio-Lens (Saha et al., 2023b)	3202	2058	786
Proposed Dataset	1950	1849	2088
Total	5152	3907	2874

Table 1: Overview of the dataset and class distribution

The dataset comprises 11,933 samples, including 6,024 samples from the Vio-Lens dataset by (Saha et al., 2023b) and 5,909 newly collected samples from platforms such as YouTube, Facebook, and TikTok. This expansion allows the model to better handle linguistic variations, including slang, spelling errors, code-mixing, and platform-specific expressions, enhancing its robustness to real-world, diverse online discourse. All data were sourced from publicly available posts, ensuring no personally identifiable information was retained. The categories are labeled as 0 for Non-Violence, 1 for Passive Violence, and 2 for Active Violence. The dataset's volume and class distribution are summarized in Table 1, where NV, PV, and AV represent Non-Violence, Passive Violence, and Active Violence, respectively. Notably, while the Vio-Lens dataset exhibits significant class imbalance, with 3,202 Non-Violence samples, 2,058 Passive Violence samples, and 786 Active Violence samples, our proposed dataset addresses this by balancing the categories. For example, Non-Violence samples in our dataset are reduced to 1,950, Passive Violence to 1,849, and Active Violence increased to 2,088. This balancing improves the model's robustness, scalability, and generalization across all violence types. The following figure 2 demonstrates the overall dataset ratio among the three labels of our dataset.

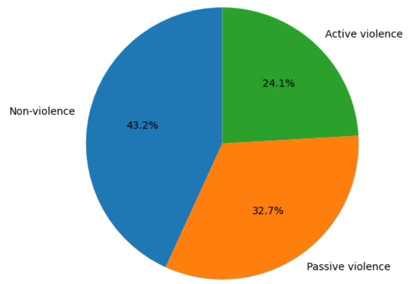

Figure 2: Class distribution of the combined dataset

The Venn diagram 3 shows the overlap and class-specific distribution of the most frequent words across the three violence categories, revealing both shared vocabulary and distinctive linguistic cues.

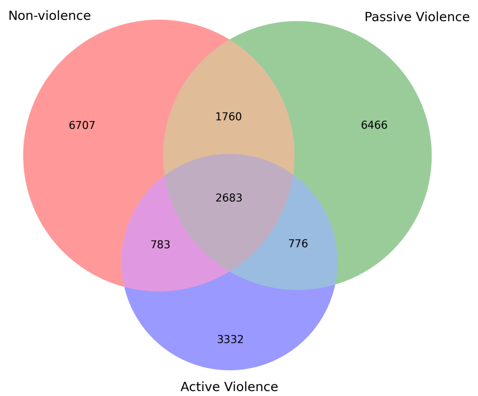

Figure 3: Venn Diagram of Top Frequent Words Across Violence Classes

Text Length Statistics: Table 2 presents the text length statistics, with a minimum of 3 characters, a maximum of 640, an average of 90.49, and a standard deviation of 78.54, indicating significant variability in text length. This variation is crucial for assessing the robustness of text-based models.

Table 2: Statistical Summary of Text Lengths

Metric	Values
Minimum Text Length	3
Maximum Text Length	640
Mean Text Length	90.49
Standard Deviation	78.54

Text Length Distribution Analysis: Figure 4 shows a right-skewed distribution of text lengths, with most texts between 20-150 characters, peaking around 40-60 characters. The long tail extending to 600+ characters suggests occasional longer texts. This distribution highlights the prevalence of short-form content, impacting model selection and preprocessing in NLP tasks.

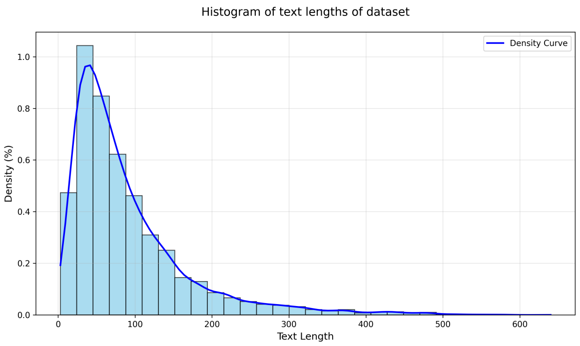

Figure 4: Histogram of test lengths of our dataset

3.2 Existing Dataset

The foundation of our dataset is the Vio-Lens dataset by (Saha et al., 2023b), which comprises 6,046 expert-annotated YouTube comments related to communal violence in Bangladesh and West Bengal. These comments were sourced from videos corresponding to nine major historical incidents between 2012 and 2022 (e.g., the Ramu violence and the 2021 Cumilla Durga Puja clashes). The dataset employs a three-class labeling scheme *Non-Violence*, *Passive Violence*, and *Active Violence*.

3.3 New Data Collection

We manually collected 5,909 additional Bangla samples from YouTube, Facebook, and TikTok, focusing on incidents from 2024 and 2025 that are likely to provoke discussions on violence, peace, or communal tensions. This collection ensures both temporal relevance and linguistic diversity, covering a range of socially and politically sensitive events, such as student protests, political developments, communal incidents, high-profile criminal cases, and sports-related news. All collected texts were filtered to retain Bangla-dominant content, duplicates were removed, and the remaining samples were reviewed for quality. Table 3 provides a summary of the platforms and representative events included in the newly collected samples.

Table 3: Platforms and Key Events for Newly Collected Samples

Platform	Event or Incident
YouTube	2024 Student Quota Reform Protest; Alleged disrespect to the Holy Quran by Apurbo Pal; Release of Zakir Khan; News related to Obaydul Quader; Political developments of Rumin Farhana; Reports on rape and murder cases; Bangladesh cricket team news.
Facebook	Jamuna TV and other national news pages; Khagrachari rape case; Sakib Al Hasan Facebook post reactions; Political news; Alleged disrespect to the Holy Quran by Apurbo Pal; Public reactions to Dr. Sabrina incident.
TikTok	A recent viral phenomenon related to the above incidents.

Figure 5 shows the dataset collection process from YouTube, Facebook, and TikTok, resulting in 11,933 samples.

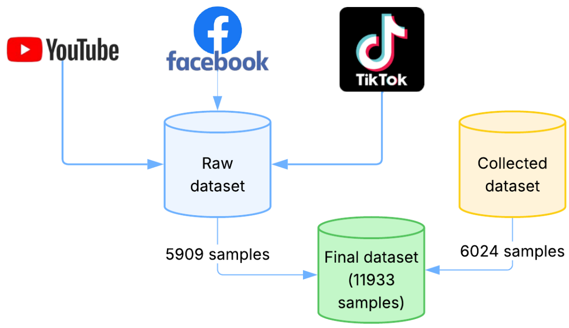

Figure 5: Overview of the dataset collection process

3.4 Data Preprocessing

The combined dataset comprised 11,945 samples from multiple sources. After removing invalid labels (9.0) and missing values, 11,933 samples remained, distributed across three categories: Non-violence (0.0, 43.2%), Passive violence (1.0, 32.7%), and Active violence (2.0, 24.1%). Preprocessing included the following steps:

3.4.1 Emoji and Noise Removal

A regular-expression-based function was applied to remove emojis and informal symbols while preserving Bengali Unicode characters, digits, and essential punctuation. This ensured that non-linguistic artifacts did not interfere with text analysis.

3.4.2 Tokenization and Corpus Construction

Text was tokenized using a Bangla-aware tokenizer, yielding a corpus of 212,488 tokens. This tokenized corpus formed the basis for vocabulary analysis and feature extraction.

3.4.3 Stopwords Filtering

Stopwords from both the NLTK Bengali corpus and BNLP were used. The BNLP stopword list was augmented with additional high-frequency function words, including *kore*, *ei*, *ki*, *ar*, *rer*, *i*, *saathe*, and *kotha*. Frequency analysis revealed the most common stopwords, such as *na* (3,231 occurrences), *kore* (2,168), and *ei* (1,875), which were filtered during cleaning.

3.4.4 Text Cleaning Pipeline

The cleaned text was generated through sequential operations: removal of Bengali punctuation, newline and whitespace anomalies, digits, Bengali and English stopwords, ASCII punctuation, and emojis. The sanitized text was stored in a new column, preserving the original content for reference.

3.4.5 Vocabulary Analysis

High-frequency lexical items were extracted from the cleaned corpus. Frequent terms included *bichar* (justice), *Allah* (Allah), *bhai* (brother), *Muslim* (Muslim), *hijab* (hijab), *shikkha* (education), and *dhormo* (religion), reflecting sociocultural and religious themes prevalent in Bengali violence-related discourse.

This preprocessing pipeline ensured thorough cleaning while preserving the semantic content necessary for robust three-class violence classification.

4 Methodology

Figure 6 presents the overall process architecture of the proposed methodology. The workflow starts with dataset acquisition from primary data repositories, followed by data integration and preprocessing. The processed dataset is then split into training, validation, and testing sets (70:15:15) for model implementation using both traditional machine learning and transformer-based approaches. Finally, model performance is evaluated and compared through different metrics and visual analysis.

This study investigates violence-inciting Bangla text classification using three categories of models: linguistic features with traditional machine learning approaches, transformer-based deep learning models, and our proposed novel BX stacking ensemble model. A unified experimental setup was adopted to ensure fair comparison across models.

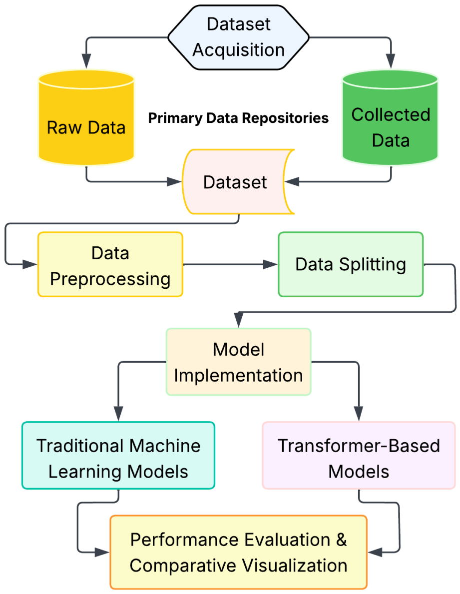

Figure 6: Overall Process Architecture of the Proposed Methodology

4.1 Linguistic Features with Traditional Machine Learning Models

We extracted lexical linguistic features from Bangla text using TF-IDF, focusing on word-level n-grams (unigrams, bigrams, and trigrams) and character-level n-grams (with lengths 3, 4, and 5). These features were designed to capture lexical and subword-level information relevant to violence-related expressions. For classification, we used Logistic Regression (LR) and Support Vector Machine (SVM), which are effective for handling high-dimensional sparse feature spaces. We also evaluated the impact of combined feature sets, such as unigram+bigram+trigram and concatenated character n-grams (c3+c4+c5), on model performance. To enhance the model's semantic understanding, we incorporated pre-trained word embeddings, capturing deeper semantic information beyond surface-level features.

4.2 Transformer-Based Models

To model contextual and semantic dependencies in Bangla text, several transformer-based architectures were evaluated. The dataset was split using stratified sampling. Text inputs were tokenized with model-specific tokenizers, and sequence lengths were adjusted accordingly. To address class imbalance, weighted loss functions were used. Five pre-trained models, such as BanglaBERT, SagorBERT, multilingual BERT (cased and uncased), and XLM-RoBERTa, were fine-tuned using the AdamW optimizer, with hyperparameter optimization for SagorBERT via Optuna. Gradient clipping and mixed-precision training were applied for stability and efficiency.

4.3 Stacking Ensemble Model

In this study, we use the BX Stacking Ensemble Model, which combines the strengths of two pre-trained transformer models: BanglaBERT and XLM-RoBERTa. BanglaBERT is tailored for Bangla text, while XLM-RoBERTa handles multilingual inputs. The model generates feature embeddings from the [CLS] token of both models, which are then concatenated into a 1536-dimensional vector. This combined feature vector is passed through a dropout layer and then through a linear classifier that serves as a meta-model, learning to combine the features from both models for final classification. By stacking the features, the model leverages both language-specific and multilingual knowledge, improving performance on diverse text inputs. This approach allows the classifier to adaptively combine the strengths of each model, enhancing the overall classification accuracy. To better understand the BX Stacking Ensemble Model, consider the architecture diagram in Figure 7, which illustrates how the model integrates the outputs from BanglaBERT and XLM-RoBERTa. This visualization captures the flow of data through the feature extraction, concatenation, dropout, and final classification stages.

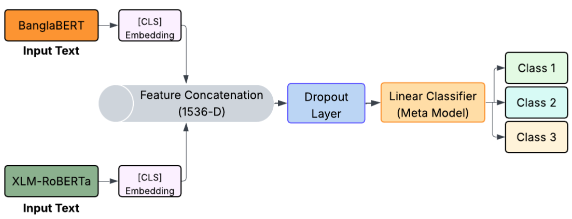

Figure 7: Model Architecture of Our Proposed BX Ensemble Model

Figure 7 depicts the BX Stacking Ensemble Model, which combines BanglaBERT and XLM-RoBERTa to improve classification accuracy by leveraging both language-specific and multilingual knowledge.

5 Experimental Setup

We evaluated traditional machine learning models (SVM and LR) and transformer-based models for classifying the dataset. Machine learning models were trained using unigram, bigram, trigram, and character-level n-grams and evaluated based on precision, recall, F1-score, and accuracy. Feature combinations like U+B+T were also tested to assess performance. For transformer-based models, we fine-tuned several architectures, including custom transformers, LSTM, BiLSTM, and SSD MobileNet V2, with a learning rate of 2e-5 and trained for 50 epochs using early stopping based on Macro F1-score. Data augmentation was applied, and performance was mainly evaluated on accuracy. The BX Stacking Ensemble Model combined the strengths of BanglaBERT and XLM-RoBERTa, with feature embeddings concatenated from both models' [CLS] tokens. The combined vector passed through a dropout layer and a linear classifier. The model was trained for 5 epochs with a batch size of 16 and a learning rate of 2e-5, addressing class imbalance through class weights. The model was evaluated on classwise accuracy, F1-score, precision, and recall. All resources are publicly available in this repository[1].

6 Result and Analysis

This section presents the evaluation and comparison of traditional machine learning models, transformer-based models, and the BX Stacking Ensemble Model, based on the F1 Score and overall macro-F1 score, to justify the class imbalance.

6.1 Evaluation metrics

The performance of all models was evaluated using standard classification metrics, including macro-averaged and weighted-averaged precision, recall, and F1-score. Macro-averaged metrics assess model performance equally across all classes, while weighted-averaged metrics account for class imbalance by weighting each class according to its support. This evaluation strategy ensures a fair and reliable comparison across different learning approaches.

6.2 Performance Comparison

To evaluate the performance of the models, we compared several approaches using the dataset, which consists of multiple categories related to violence detection. The models were assessed on Non-Violence, Passive Violence, Active Violence, and the overall Macro F1 score. The following table 4 presents the results for both traditional ma-

[1] Github:https://anonymous.4open.science/r/ACL-2026/README.md

chine learning models and transformer-based models, including the BX Stacking Ensemble Model. These models were evaluated using F1 scores for each category, providing a comprehensive view of their classification performance.

Table 4: Model Performance Comparison

Models	F1			
	NV	PV	AV	Macro
Linguistic features with SVM and LR				
Unigram(U)+SVM	0.77	0.69	0.7	0.72
Bigram(B)+LR	0.63	0.34	0.41	0.46
Trigram(T)+LR	0.3	0.61	0.01	0.26
(U+B+T)+SVM	0.78	0.7	0.7	0.73
C3-Gram(C3)+SVM	0.8	0.73	0.73	0.75
C4-Gram(C4)+SVM	0.8	0.74	0.76	0.77
C5-Gram(C5)+SVM	0.81	0.74	0.78	0.78
(C3+C4+C5)+SVM	0.81	0.75	0.77	0.78
BERT Models				
SagorBERT	0.81	0.74	0.79	0.78
BanglaBERT	0.83	0.77	0.81	0.8
M-BERT-Cased	0.8	0.75	0.77	0.77
M-BERT-unCased	0.82	0.75	0.78	0.79
XLMRoBERTa	0.82	0.76	0.8	0.79
Stacking Ensemble Model				
BX Ensemble (proposed)	**0.9**	**0.88**	**0.78**	**0.85**

As shown in Table 4, the BX Stacking Ensemble Model outperforms individual models, achieving the highest F1 scores across all categories, including Non-Violence, Passive Violence, Active Violence, and Macro F1. From the Linguistic Features with Machine Learning Models section, the combination of C3+C4+C5 with SVM achieves the highest Macro F1 score of 0.78. Among the Transformer-based models, BanglaBERT performs best with a Macro F1 score of 0.80. Finally, our BX Ensemble Model surpasses these models, achieving the highest performance with a Macro F1 score of 0.85, demonstrating the effectiveness of combining the strengths of both BanglaBERT and XLM-RoBERTa in the ensemble approach.

6.3 Visualization of models comparison

Figure 8 presents a bar plot comparing the Macro F1 scores of 14 models evaluated in this study, grouped into three categories: Linguistic Features with Machine Learning Models, BERT Models, and the BX Stacking Ensemble Model. The plot clearly highlights the relative performance of these models, emphasizing the strength of the BX Stacking Ensemble Model.

Figure 8 shows that the BX Stacking Ensemble Model achieves the highest Macro F1 score of 0.85, outperforming traditional machine learning and individual transformer models, demonstrating its effectiveness in improving performance across all violence-related categories.

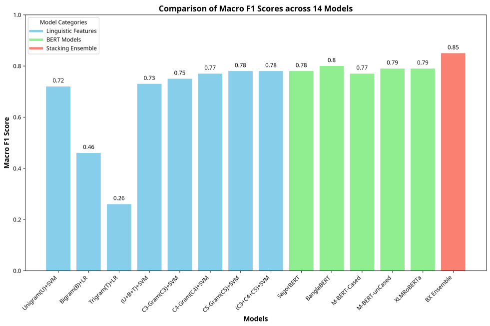

Figure 8: Comparison of Macro F1 Scores Across 14 Models

6.4 Performance Improvement and Effect Size Analysis

The BX Ensemble model achieves the highest Macro F1 score of 0.85, surpassing the best baseline, BanglaBERT (0.80), by 6.25%. Effect size analysis shows large practical significance with Cohen's d values of 2.0-2.4, indicating substantial improvements. Class-wise evaluation reveals BX Ensemble excels in Non-Violence (F1: 0.90, +8.43%) and Passive Violence (F1: 0.88, +14.29%), with a 100% win rate against all baselines. While BanglaBERT slightly outperforms on Active Violence (0.81 vs 0.78), BX Ensemble maintains a 61.5% win rate. This validates the effectiveness of the ensemble approach in leveraging complementary model strengths across all violence types.

7 Conclusion

This study introduces an automated system for detecting violence-inciting Bangla text using a BX Stacking Ensemble Model, combining BanglaBERT and XLM-RoBERTa. Trained on a diverse dataset of 11,933 samples, including the Vio-Lens dataset and 5,909 additional instances from social media platforms, the model captures various linguistic nuances, enhancing accuracy across all violence-related categories. By integrating both language-specific and multilingual models, it achieves improved robustness and adaptability, outperforming individual models. This research highlights the importance of diverse model architectures and provides a scalable solution to combat harmful content on social media, especially in low-resource languages like Bangla.

8 Limitations

This study faces limitations, including the high computational resource requirements of transformer models, which can be challenging for real-time deployment. Additionally, there is a lack of sufficient labeled data in publicly available Bangla datasets, which limits the model's ability to generalize effectively. Future work can address these limitations by expanding datasets through data augmentation and exploring lighter transformer models. Incorporating transfer learning and enhancing contextual understanding can further improve the model's accuracy and adaptability.

References

Md Tofael Ahmed, Almas Antar, Maqsudur Rahman, A. Islam, Dipankar Das, and Md. Golam Rashed. 2023. Social media cyberbullying detection on political violence from bangla texts using machine learning algorithm. *Journal of Intelligent Learning Systems and Applications*, 15:108–122.

Refaat Mohammad Alamgir and Amira Haque. 2023. Team CentreBack at BLP-2023 task 1: Analyzing performance of different machine-learning based methods for detecting violence-inciting texts in Bangla. In *Proceedings of the First Workshop on Bangla Language Processing (BLP-2023)*, pages 168–173, Singapore. Association for Computational Linguistics.

Udoy Das, Karnis Fatema, Md Ayon Mia, Mahshar Yahan, Md Sajidul Mowla, Md Fayez Ullah, Arpita Sarker, and Hasan Murad. 2023. EmptyMind at BLP-2023 task 1: A transformer-based hierarchical-BERT model for Bangla violence-inciting text detection. In *Proceedings of the First Workshop on Bangla Language Processing (BLP-2023)*, pages 174–178, Singapore. Association for Computational Linguistics.

Krishno Dey, Prerona Tarannum, Md. Arid Hasan, and Francis Palma. 2023. Semantics squad at BLP-2023 task 1: Violence inciting Bangla text detection with fine-tuned transformer-based models. In *Proceedings of the First Workshop on Bangla Language Processing (BLP-2023)*, pages 225–229, Singapore. Association for Computational Linguistics.

Joseph Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76:378–382.

Abdullah Khondoker, Enam Ahmed Taufik, Md Iftekhar Islam Tashik, SM Mahmud, and Farig Sadeque. 2025. How effectively can bert models interpret context and detect bengali communal violent text? *arXiv preprint arXiv:2506.19831*.

Sourabrata Mukherjee, Atul Kr Ojha, and Ondřej Dušek. 2023. Ufal-uld at blp-2023 task 1: Violence detection in bangla text. In *Proceedings of the First Workshop on Bangla Language Processing (BLP-2023)*, pages 220–224.

Saurabh Page, Sudeep Mangalvedhekar, Kshitij Deshpande, Tanmay Chavan, and Sheetal Sonawane. 2023. Mavericks at blp-2023 task 1: Ensemble-based approach using language models for violence inciting text detection. *Preprint*, arXiv:2311.18778.

Md Nishat Raihan, Dhiman Goswami, Sadiya Sayara Chowdhury Puspo, and Marcos Zampieri. 2023. nlpBDpatriots at BLP-2023 task 1: Two-step classification for violence inciting text detection in Bangla - leveraging back-translation and multilinguality. In *Proceedings of the First Workshop on Bangla Language Processing (BLP-2023)*, pages 179–184, Singapore. Association for Computational Linguistics.

Omar Faruqe Riyad, Trina Chakraborty, and Abhishek Dey. 2023. Team_Syrax at BLP-2023 task 1: Data augmentation and ensemble based approach for violence inciting text detection in Bangla. In *Proceedings of the First Workshop on Bangla Language Processing (BLP-2023)*, pages 247–254, Singapore. Association for Computational Linguistics.

Saumajit Saha and Albert Nanda. 2023. Banglanlp at blp-2023 task 1: Benchmarking different transformer models for violence inciting text detection in bengali. *arXiv preprint arXiv:2310.10781*.

Sourav Saha, Jahedul Alam Junaed, Maryam Saleki, Mohamed Rahouti, Nabeel Mohammed, and Mohammad Ruhul Amin. 2023a. BLP-2023 task 1: Violence inciting text detection (VITD). In *Proceedings of the First Workshop on Bangla Language Processing (BLP-2023)*, pages 255–265, Singapore. Association for Computational Linguistics.

Sourav Saha, Jahedul Alam Junaed, Maryam Saleki, Arnab Sen Sharma, Mohammad Rashidujjaman Rifat, Mohamed Rahouti, Syed Ishtiaque Ahmed, Nabeel Mohammed, and Mohammad Ruhul Amin. 2023b. Vio-lens: A novel dataset of annotated social network posts leading to different forms of communal violence and its evaluation. In *Proceedings of the First Workshop on Bangla Language Processing (BLP-2023)*, pages 72–84, Singapore. Association for Computational Linguistics.

Hrithik Shibu, Shrestha Datta, Zhalok Rahman, Shahrab Sami, Md. Sumon Miah, Raisa Fairooz, and Md Mollah. 2023. SUST_Black box at BLP-2023 task 1: Detecting communal violence in texts: An exploration of MLM and weighted ensemble techniques. In *Proceedings of the First Workshop on Bangla Language Processing (BLP-2023)*, pages 208–213, Singapore. Association for Computational Linguistics.

Nazia Tasnim, Sujan Sen Gupta, Md Istiak Hossain Shihab, Fatiha Islam Juee, Arunima Tahsin, Pritom Ghum, Kanij Fatema, Marshia Haque, Wasema Farzana, Prionti Nasir, and 1 others. 2024. Mapping violence: Developing an extensive framework to build a bangla sectarian expression dataset from social media interactions. *arXiv preprint arXiv:2404.11752*.