Velora at BLP-2025 Task 1: Multi-Method Evaluation for Hate Speech Classification in Bangla Text

Sad Yeamin Sayem
BRAC University
Department of CSE
sad.yeamin.sayem@g.bracu.ac.bd

Sabira Rahman
BRAC University
Department of CSE
sabira.rahman1@g.bracu.ac.bd

Abstract

Hate speech detection in Bangla is challenging due to complex morphology, frequent code-mixing, and severe class imbalance across categories such as abuse, sexism, religious and political hate, profanity, and neutrality. The BLP Workshop 2025 Subtask 1A addressed this by classifying Bangla YouTube comments into these categories to support online moderation in low-resource settings. We developed a BanglaBERT-based system with balanced data augmentation and advanced regularization techniques, combined with optimized training strategies for better generalization. On the blind test set, our system achieved a micro F1 score of 0.7013, ranking 21st on the leaderboard. These results indicate that augmentation, robust loss functions, and model refinements can enhance Bangla hate speech detection, though implicit and context-dependent hate speech remains difficult.

1 Introduction

Hate speech detection is the task of automatically identifying harmful or offensive language directed towards individuals or groups. Hate speech can take many forms, including derogatory remarks based on race, religion, gender, politics, nationality, or the use of profanities and abusive expressions. Such content poses risks by reinforcing social inequalities, deepening divides, and in some cases fueling real-world hostility. With the rapid growth of online communication, automated hate speech detection has become an important area of research to support moderation and promote safer digital spaces. Recent work has explored similar challenges in other languages, such as the development of a Slovak hate-speech detection system using a neural-network-based approach applied to comments on social media (Sokolová et al., 2022). However, beyond explicit slurs, many instances of hate speech are context-dependent or subtle. These include sarcastic remarks, coded expressions such as references to 'those people' or 'your kind', indirect political or religious insinuations, and statements whose hateful intent becomes evident only through a conversational or cultural context (Ocampo et al., 2023).

In the context of Bangla, these challenges are amplified. Bangla is the seventh most spoken language in the world, with over 230 million speakers, and online content in Bangla is expanding rapidly (Momin and Sarker, 2025a). However, effective moderation tools tailored for Bangla are still limited. The language itself is morphologically rich (Faridee and Tyers, 2009), frequently mixes with English in social media, and exhibits severe class imbalance in hateful expressions across different categories (Chanda et al., 2016). Research on Bangla hate speech detection began only recently; earlier work relied on traditional machine learning methods, followed by deep learning approaches such as CNNs and LSTMs (Rawal and Asirvatham, 2025; Kumar et al., 2024). More recently, transformer-based models like mBERT (Nozza et al., 2020), XLM-RoBERTa (Singh et al., 2023), and BanglaBERT (Bhattacharjee et al., 2021) have demonstrated promising results, supported by shared tasks such as HASOC (Mandl et al., 2025) and dedicated hate speech benchmarks.

Building on this foundation, our work contributes in three key ways:

- **Balanced data augmentation:** combining undersampling of frequent categories with noisy oversampling of rare ones.

- **Model refinements:** adding techniques like mean pooling, multi-sample dropout, and a linear classifier head, trained with Class-Balanced Focal Loss and R-Drop regularization.

Proceedings of the Second Workshop on Bangla Language Processing (BLP-2025), pages 531–537
December 23, 2025 ©2025 Association for Computational Linguistics

- **Advanced training strategies:** including layer-wise learning rate decay, cosine warmup, gradient checkpointing, mixed-precision training, gradient accumulation, and early stopping.

With these improvements, we achieved a micro-F1 score of **0.7013** on BLP 2025 Subtask 1A. This suggests that combining balanced data augmentation with targeted model refinements can yield measurable improvements in hate speech detection in Bangla, although context-dependent and subtle cases remain difficult. Our code and experimental setup are publicly available to facilitate further research in Bangla hate speech detection[1].

2 Related Work

Hate speech detection has been extensively studied in high-resource languages such as English, Hindi, Urdu, Arabic, etc.(Usman et al., 2025; Jahnavi and Chaturvedi, 2025; Ahmad et al., 2024) Early approaches relied on handcrafted features and traditional classifiers such as Naive Bayes, SVM, and Random Forests(Mullah and Zainon, 2021). While these methods demonstrated the feasibility of automated detection, their inability to capture deeper semantics limited generalization (Pruengkarn et al., 2025). The advent of pre-trained language models, such as BERT and its multilingual variants (Nozza et al., 2020), introduced contextual embeddings that became the dominant approach, offering substantial improvements in robustness and cross-lingual transfer (Papel et al., 2024).

In the Bangla context, systematic research is comparatively recent. Initial efforts focused on word embeddings like Word2Vec, FastText (Arif et al., 2024), and GloVe (Mahmud et al., 2022) in combination with classifiers, which achieved moderate success but struggled with challenges unique to Bangla, including morphological richness, frequent code-mixing with English, and severe class imbalance (Momin and Sarker, 2025b).

Several benchmark data sets have been released to support progress in detecting hate speech in Bangla. BanglaBERT (Bhattacharjee et al., 2021) introduced a dataset covering abusive and offensive content, while the Hate Speech and Offensive Content (HASOC) shared tasks (Mandl et al., 2025) included Bangla as one of the languages, providing labeled data for multilingual evaluation. More recently, BIDWESH (Fayaz et al., 2025) expanded its coverage to dialectal and informal Bangla, broadening the scope of evaluation. The Bangla Language Processing (BLP) Workshop 2025 (Hasan et al., 2025b) provided one of the largest curated hate speech benchmarks to date, standardizing evaluation and encouraging cross-system comparisons. Despite these advances, challenges such as low-resource categories (e.g., sexism, religious hate) and context-dependent expressions remain largely unresolved, motivating further exploration of balanced augmentation and model refinements.

3 Task & DataSet Overview

3.1 Task Overview

The Bangla Multi-task Hate Speech Identification shared task (Hasan et al., 2025a), part of the BLP Workshop 2025, aimed to classify Bangla YouTube comments into six categories as abuse, sexism, religious hate, political hate, profanity, and neutral, to establish a benchmark for developing robust hate speech detection systems in Bangla, where challenges such as code-mixing, complex morphology, and severe class imbalance make moderation particularly difficult. Subtask 1A evaluated the systems on this single-label classification task under realistic low-resource conditions.

3.2 Dataset Description

We constructed the training data set by merging the dataset provided for Subtask 1A of the BLP Workshop 2025 (Hasan et al., 2025b) with another publicly available dataset (Karim et al., 2020) which we selected because its categories align well with the BLP Workshop dataset, allowing for a seamless merge and a more comprehensive training set. To ensure consistency, label categories were normalized using a mapping strategy, where synonymous or overlapping labels (e.g., Political and Geopolitical) were merged under a unified class (Political Hate), and ambiguous labels were refined (e.g., Personal → Abusive, Gender abusive → Sexism). After cleaning and harmonization, the final dataset contains six classes: None (neutral/non-hateful), Abusive, Political Hate, Profane, Religious Hate, and Sexism. The class distribution is highly imbalanced, with 19,955 neutral samples. This imbalance reflects the real-world prevalence of hate categories in Bangla social media and motivates the use of augmentation and loss re-weighting strate-

[1] https://github.com/programophile/Bangla_NLP_workshop_Subtask1A_Velora

gies during training.

The merged training set consists of **38,941 samples** (35,523 from shared & 3,418 from the public dataset). Meanwhile, the blind development test set (`dev_test`) contains **2,512** unlabeled samples reserved for evaluation. Table 2 presents the final class distribution of the merged training set.

4 Methodology

4.1 Models & Workflow

All five models were based on BanglaBERT, with variations introduced in architecture, loss functions, and data augmentation. The main challenge across them was severe class imbalance, with categories such as `Sexism` and `Religious Hate` remaining underrepresented and difficult to classify, consistently limiting performance.

Figure 1 outlines the entire pipeline: the data set is balanced and augmented, tokenized, and encoded with BanglaBERT to produce contextual embeddings. These embeddings undergo mean pooling and multi-sample dropout before classification through a linear head trained with R-Drop and class-balanced focal loss to improve robustness.

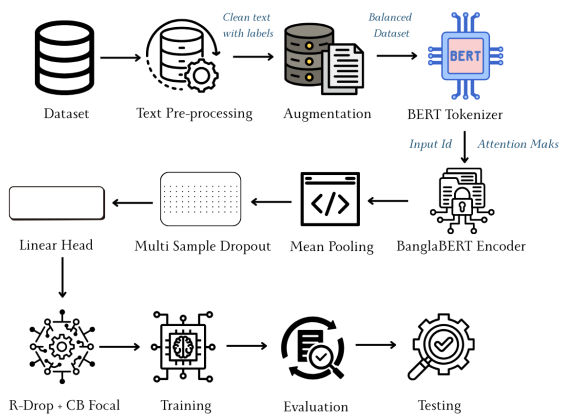

Figure 1: Workflow of the proposed Bangla hate speech classifier: Raw Text → Preprocessing (cleaning & normalization) → Data Augmentation (mitigating class imbalance) → BanglaBERT Encoding (contextual embeddings) → Mean Pooling & Multi-Sample Dropout (generalization) → Classification Head (R-Drop + Class-Balanced Focal Loss), producing robust and fair predictions across all hate speech categories.

4.1.1 Advanced BanglaBERT Fine-tune with CB-Focal & R-Drop

Our submitted approach (Advanced BanglaBERT Fine-tune with CB-Focal & R-Drop) used BanglaBERT as the backbone, leveraging its transformer-based contextual embeddings shown in previous studies to effectively generalize across various Bangla NLP tasks as a strong foundation for classification (Kowsher et al., 2022). The input data was normalized through targeted text cleaning to reduce noise from informal YouTube comments. To address class imbalance, we capped the dominant “None” class via undersampling and duplicated minority class samples with light noisy augmentations. This balanced training exposure prevented the model from collapsing into majority predictions. At the architecture level, we applied mean pooling across token embeddings for efficient sentence-level representation, followed by a single linear classifier head. Generalization was encouraged through multi-sample dropout, which averages predictions across multiple dropout masks, and R-Drop, which enforces consistency across perturbed forward passes. Training optimization combined Class-Balanced Focal Loss, which emphasized difficult minority examples, with layer-wise learning rate decay and cosine scheduling for stable fine-tuning. Training used gradient checkpointing, FP16 mixed precision, gradient accumulation, and early stopping to manage memory usage and stabilize the optimization process.

4.1.2 Balanced Augmented BanglaBERT

Balanced Augmented BanglaBERT introduced two main changes while following the same overall setup as Advanced BanglaBERT Fine-tune with CB-Focal & R-Drop. For data balancing, we replaced simple duplication with controlled duplication combined with word shuffling, which increased sample diversity without redundancy. Inside the model, we replaced the single linear head with a lightweight two-layer MLP, improving convergence stability and reducing overfitting. All other training strategies followed the previous system. These modifications yielded a slightly better result.

4.1.3 BanglaBERT-Hybrid (CNN-BiLSTM-Attn)

This model used standard BanglaBERT fine-tuning with cross-entropy loss. The lack of balancing or augmentation caused a strong bias toward the dominant `None` class and low recall in minority categories. While surface-level hate speech was often detected, context-dependent instances were frequently missed. It achieved a micro-F1 of 0.6954, but without class balancing, it overfit the majority class and showed poor recall for minority labels, resulting in many subtle or contextual hate expressions being misclassified.

4.1.4 Conservative BanglaBERT-Hybrid

This variant extended BanglaBERT with a hybrid CNN-BiLSTM-Attention architecture to capture both local n-gram patterns and long-range dependencies. Training with Class-Balanced Focal Loss improved recall for minority classes, but conservative preprocessing and limited augmentation restricted data diversity, causing underfitting in rare categories. The model achieved a micro-F1 of 0.6793, showing modest but consistent gains over the base system. While minority-class recall improved slightly, precision remained stable and overall recall was still constrained.

4.1.5 Optimized BanglaBERT-Hybrid

This model introduced back-translation, a data augmentation method in which Bangla comments were translated into English and then back into Bangla to generate paraphrased variants that preserved meaning while increasing linguistic diversity. It also applied logit-adjusted cross-entropy to mitigate class imbalance, R-Drop for regularization, and snapshot ensembling for stable predictions. These refinements yielded modest gains and better robustness, though performance remained limited by noisy data and the long-tail label distribution.

4.2 Types of Augmentations Performed on the Minority Classes

Across all variants of the model, the augmentation was applied exclusively to minority hate-speech classes to mitigate the severe class imbalance. The **BanglaBERT-Hybrid (CNN-BiLSTM-Attn)** and **Conservative BanglaBERT-Hybrid** models used light oversampling through direct duplication and controlled word-order shuffling to introduce limited lexical variation. The **Balanced Augmented BanglaBERT** and **Optimized BanglaBERT-Hybrid** systems incorporated stronger augmentation, including targeted oversampling, local word shuffling, and back-translation, particularly for underrepresented categories such as religious hate and sexism, to create semantically diverse paraphrases. The **Advanced BanglaBERT Fine-tune submitted with CB-Focal & R-Drop** applied only light noise-based duplication while capping the dominant None class to maintain balance. Overall, augmentation evolved from simple duplication to richer semantic transformations, consistently targeting minority labels.

5 Results & Findings

The task was evaluated using micro-F1, which balances performance across all classes. As shown in Table 1, our primary model **Advanced BanglaBERT with CB-Focal & R-Drop** achieved 0.7013. The best-performing system, **Balanced Augmented BanglaBERT**, improved this to 0.7025 through controlled augmentation and a lightweight classifier, showing that careful balancing strategies provide more benefit than architectural complexity.

Hybrid models such as **BanglaBERT-Hybrid** and the **Optimized Hybrid** underperformed compared to simpler transformer-based setups, while the **Conservative Hybrid** was lowest. Overall, results highlight that balanced augmentation yields modest but consistent improvements, while rare-class scarcity remains the main bottleneck.

Models	**F1-Score**
`Balanced Augmented BanglaBERT`	0.7025
`Advanced BanglaBERT`	0.7013
`BanglaBERT-Hybrid`	0.6954
`Optimized BanglaBERT-Hybrid`	0.6886
`Conservative BanglaBERT-Hybrid`	0.6793

Table 1: micro-F1 scores of different systems

5.1 Category-wise Performance

To understand model behavior across individual classes, we present a comprehensive performance table (Table 3) for the best model **Balanced Augmented BanglaBERT**. These metrics are computed on the validation set used for model selection, and serve to illustrate strengths and weaknesses per category. Final evaluation results reported in Table 1 are based on the test set.

5.2 Error Analysis

Although **Advanced BanglaBERT Fine-tune with CB-Focal & R-Drop**, incorporated strong embeddings, tailored loss functions, and advanced optimization, its effectiveness was limited. Oversampling by simple duplication reduced training diversity and encouraged overfitting. Most critically, categories such as sexism and religious hate remained severely underrepresented, causing low recall despite reweighted loss functions. These factors constrained generalization, leaving implicit or context-dependent hate speech difficult to detect.

In **Balanced Augmented BanglaBERT**, improvements over the submitted system were small. The more balanced augmentation and simplified architecture increased diversity modestly, but rare classes still lacked sufficient coverage. As a result, recall for underrepresented categories improved slightly, yet overall performance remained limited by data scarcity and linguistic complexity.

We also examined weaker variants to understand performance limitations. The **BanglaBERT-Hybrid (CNN-BiLSTM-Attn)** with micro-F1 0.6954, employed a complex hybrid architecture combining CNN, BiLSTM, multi-head attention, and a deep classifier on top of BanglaBERT. While intended to capture richer representations, the added complexity caused overfitting and unstable training. Aggressive focal loss penalized majority classes excessively, further reducing generalization. Consequently, this model failed to surpass simpler model despite higher model capacity.

The **Conservative BanglaBERT-Hybrid** with micro-F1 0.6793 was under-engineered. Its minimal data augmentation left minority categories underrepresented. The class weighting was limited in range, insufficient to address imbalance, and the shallow CNN–LSTM layers lacked expressive power for nuanced hate speech. This led to underfitting on rare classes, making it the least performing system.

Optimized BanglaBERT-Hybrid (micro-F1 0.6886) added back-translation augmentation, logit-adjusted loss, and snapshot ensembling. Despite these improvements, performance gains were just adequate. Back-translation often produced semantically similar sentences rather than truly diverse ones, limiting benefits for rare categories such as sexism and religious hate. Logit-adjusted loss and ensembling stabilized training and slightly improved recall, but severe underrepresentation remained, causing low class-wise F1 for rare labels. Implicit, context-dependent, or subtle hate speech continued to be misclassified, reflecting the reliance on shallow augmentations and surface-level lexical cues. These results indicate that incremental improvements from loss adjustment and ensembling are insufficient without richer data and stronger coverage of minority classes.

6 Conclusion

In this work, we presented multiple Bangla hate speech detection systems for the BLP 2025 shared task 1A. Our best model, **Balanced Augmented BanglaBERT**, achieved a micro-F1 of 0.7025, demonstrating that careful augmentation and lightweight architectures can outperform more complex hybrids. Persistent imbalance in rare classes continues to cap performance, highlighting the need for richer datasets and context-aware augmentation in future work.

Limitations

Our experiments were conducted under compute and time constraints, which restricted exhaustive hyperparameter tuning and limited the exploration of more advanced augmentation strategies. The dataset, while harmonized from existing resources, remained highly imbalanced, with rare categories such as sexism and religious hate severely underrepresented. Augmentation methods like duplication and back-translation produced limited diversity, constraining generalization. Furthermore, the informal nature of Bangla YouTube comments introduced noise and inconsistencies in preprocessing, which may have affected model robustness. Finally, all systems were trained and evaluated on the shared task dataset; broader validation across larger, more diverse Bangla corpora would strengthen the generalizability of our findings.

References

Ashraf Ahmad, Mohammad Azzeh, Eman Alnagi, Qasem Abu Al-Haija, Dana Halabi, Abdullah Aref, and Yousef AbuHour. 2024. Hate speech detection in the arabic language: corpus design, construction, and evaluation. *Frontiers in Artificial Intelligence*, Volume 7 - 2024.

Md Ariful Islam Arif, Md Mahbubur Rahman, Md. Golam Rabiul Alam, and M. Akhtaruzzaman. 2024. Analyzing the performance of deep learning models for detecting hate speech on social media platforms. *MIST INTERNATIONAL JOURNAL OF SCIENCE AND TECHNOLOGY*, 12(2):29–42.

Abhik Bhattacharjee, Tahmid Hasan, Wasi Uddin Ahmad, Kazi Samin, Md Saiful Islam, Anindya Iqbal, M. Sohel Rahman, and 1 others. 2021. Banglabert: Language model pretraining and benchmarks for low-resource language understanding evaluation in bangla. *Findings of the Association for Computational Linguistics*, arXiv:2101.00204.

Arunavha Chanda, Dipankar Das, and Chandan Mazumdar. 2016. Unraveling the English-Bengali code-mixing phenomenon. In *Proceedings of the Second Workshop on Computational Approaches to Code Switching*, pages 80–89, Austin, Texas. Association for Computational Linguistics.

Abu Zaher Md Faridee and Francis M. Tyers. 2009. Development of a morphological analyser for Bengali. In *Proceedings of the First International Workshop on Free/Open-Source Rule-Based Machine Translation*, pages 43–50, Alacant, Spain.

A. H. Fayaz, M. D. Uddin, R. U. Bhuiyan, Z. Sultana, M. S. Islam, B. Paul, and S. Manzoor. 2025. BIDWESH: A bangla regional based hate speech detection dataset. *arXiv preprint arXiv:2507.16183*.

Md Arid Hasan, Firoj Alam, Md Fahad Hossain, Usman Naseem, and Syed Ishtiaque Ahmed. 2025a. Blp 2023 task 1: Hate speech identification. In *Proceedings of the Second International Workshop on Bangla Language Processing (BLP-2025)*, India. Association for Computational Linguistics.

Md Arid Hasan, Firoj Alam, Md Fahad Hossain, Usman Naseem, and Syed Ishtiaque Ahmed. 2025b. Llm-based multi-task bangla hate speech detection: Type, severity, and target. *arXiv preprint arXiv:2510.01995*.

Enduri Jahnavi and Animesh Chaturvedi. 2025. Hate speech detection in low-resource languages hindi, gujarati, marathi, and sinhala. In *2025 17th International Conference on COMmunication Systems and NETworks (COMSNETS)*, pages 7–12.

Md. Rezaul Karim, Bharathi Raja Chakravarthi, John P. McCrae, and Michael Cochez. 2020. Classification benchmarks for under-resourced bengali language based on multichannel convolutional-lstm network. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 390–399.

M. Kowsher, Abdullah As Sami, Nusrat Jahan Prottasha, Mohammad Shamsul Arefin, Pranab Kumar Dhar, and Takeshi Koshiba. 2022. Bangla-bert: Transformer-based efficient model for transfer learning and language understanding. *IEEE Access*, 10:91855–91870.

Ashwini Kumar, Santosh Kumar, Kalpdrum Passi, and Aniket Mahanti. 2024. A hybrid deep bilstm-cnn for hate speech detection in multi-social media. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 23(8).

Md. Shihab Mahmud, Md. Touhidul Islam, Afrin Jaman Bonny, Rokeya Khatun Shorna, Jasia Hossain Omi, and Md. Sadekur Rahman. 2022. Deep learning based sentiment analysis from bangla text using glove word embedding along with convolutional neural network. In *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–6.

Thomas Mandl, Koyel Ghosh, Nishat Raihan, Sandip Modha, Shrey Satapara, Tanishka Gaur, Yaashu Dave, Marcos Zampieri, and Sylvia Jaki. 2025. Overview of the hasoc track 2024: Hate-speech identification in english and bengali. In *Proceedings of the 16th Annual Meeting of the Forum for Information Retrieval Evaluation*, FIRE '24, page 1–2, New York, NY, USA. Association for Computing Machinery.

Md. Emnul Momin and Hasan Sarker. 2025a. Enhancing hate speech detection in bengali language using machine learning and deep learning algorithms. In *2025 2nd International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM)*, pages 1–6.

Md. Emnul Momin and Hasan Sarker. 2025b. Enhancing hate speech detection in bengali language using machine learning and deep learning algorithms. In *2025 2nd International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM)*, pages 1–6.

Nanlir Sallau Mullah and Wan Mohd Nazmee Wan Zainon. 2021. Advances in machine learning algorithms for hate speech detection in social media: A review. *IEEE Access*, 9:88364–88376.

Debora Nozza, Federico Bianchi, and Dirk Hovy. 2020. What the [mask]? making sense of language-specific bert models.

Nícolás Benjamí n Ocampo, Ekaterina Sviridova, Elena Cabrio, and Serena Villata. 2023. An in-depth analysis of implicit and subtle hate speech messages. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 1997–2013. Association for Computational Linguistics.

Md. Habibur Rahman Papel, Udoy Chandra Dey, and Toufiq Hasan Turza. 2024. Bangla hate speech detection by embedding and hybrid machine learning algorithms. *World Journal of Advanced Research and Reviews*. PDF at https://wjarr.com/sites/default/files/WJARR-2024-3399.pdf.

Ratchakoon Pruengkarn, Supakpong Jinarat, and Ekkasit Srisukha. 2025. Detecting sarcasm in tweets: A comparative study of deep learning and traditional approaches. In *2025 6th International Conference on Big Data Analytics and Practices (IBDAP)*, pages 165–170.

Purvi Rawal and Anisha Asirvatham. 2025. Hate speech detection using lstm and machine learning models. *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*.

Karanpreet Singh, Vajratiya Vajrobol, and Nitisha Aggarwal. 2023. IIC_Team@multimodal hate speech event detection 2023: Detection of hate speech and targets using xlm-roberta-base. In *Proceedings of the 6th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text*, pages 136–143, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Zuzana Sokolová, Ján Staš, and Daniel Hládek. 2022. An introduction to detection of hate speech and offensive language in slovak. In *2022 12th International Conference on Advanced Computer Information Technologies (ACIT)*, pages 497–501.

M. Usman, M. Ahmad, G. Sidorov, I. Gelbukh, and R. Q. Tellez. 2025. A large language model-based approach for multilingual hate speech detection on social media. *Computers*, 14(7):279.

A Appendix

Category	**Count**
None	19955
Abusive	8841
Political Hate	6198
Profane	2331
Religious Hate	1178
Sexism	438
Total	38941

Table 2: Class distribution of the merged Bangla hate speech training dataset.

	text	label
3414	এশালা খমতা ছেড়ে নির্বাচন দেনা খানকির পোলা	Personal
3415	সুচি ভারতের মুদির কুপরামর্শ নিয়ে বিবাধ সৃষ্টি করেছে	Geopolitical
3416	হাচিনার মারে আমি চুদি	Gender abusive
3417	লাভ নাই আম্লীগ আর যে কদিন ভারতও সে কদিন এমনিতেই আম্লিগ গাড়ী গোড়া বাইন্ধা রেডি আছে	Geopolitical
3418	পাকিস্তান বাংলাদেশ এক থাকলে মায়ানমার বিশাল বড় বাশ খেতো	Geopolitical

Figure 2: Sample entries from the Bangla hate-speech dataset introduced by (Karim et al., 2020). The “text” field represents original Bangla social-media posts, and the “label” field provides the manually assigned classification category used for supervised learning tasks.

Category	**Precision**	**Recall**	**F1-Score**
None	0.8356	0.7919	0.8132
Religious Hate	0.2900	0.7632	0.4203
Sexism	0.2000	0.1818	0.1905
Political Hate	0.5895	0.5773	0.5833
Profane	0.7258	0.8599	0.7872
Abusive	0.5809	0.5727	0.5768
Accuracy	–	–	**0.7189**
Macro Avg	0.5370	0.6245	0.5619
Weighted Avg	0.7320	0.7189	0.7232

Table 3: Comprehensive Performance Metrics for Advanced BanglaBERT Fine-tune with CB-Focal & R-Drop model per Category on the Validation Set.