

SyntaxMind at BLP-2025 Task 1: Leveraging Attention Fusion of CNN and GRU for Hate Speech Detection

Md. Shihab Uddin Riad

Dept. of Computer Science & Engineering
International Islamic University Chittagong
shihab.riadn@gmail.com

Abstract

This paper describes our system used in the BLP-2025 Task 1: Hate Speech Detection. We participated in Subtask 1A and Subtask 1B, addressing hate speech classification in Bangla text. Our approach employs a unified architecture that integrates BanglaBERT embeddings with multiple parallel processing branches based on GRUs and CNNs, followed by attention and dense layers for final classification. The model is designed to capture both contextual semantics and local linguistic cues, enabling robust performance across subtasks. The proposed system demonstrated high competitiveness, obtaining 0.7345 micro F1-Score (2nd place) in Subtask 1A and 0.7317 micro F1-Score (5th place) in Subtask 1B.

1 Introduction

Hate speech is any type of statement that aims to vilify, humiliate, or instigate hatred toward a group or a class of people based on their nationality, race, religion, skin color, sexual orientation, gender identity, ethnicity, or disability (Ward, 1997). With surge of various social media platforms, people often express their opinion without hesitation. However, these opinions are not always appropriate. This study aims to detect hate speech on BLP-2025 Task 1 (Hasan et al., 2025b) using the provided multiclass based dataset (Hasan et al., 2025a). BLP-2025 Task 1 consists of three different subtasks. These subtasks focus on detecting hate speech, assessing its severity, and identifying the target group of hate speech in the Bangla language. Our study participates in Subtask 1A and Subtask 1B.

Our work primarily focuses on attention based deep learning techniques to enhance representation and classification performance. As we work through the task, we notice that language specific bias in the pretrained model is significant. Due to that we have utilized BanglaBERT (Bhattacharjee et al., 2022) as contextual embedding module.

Then we pass these embeddings in CNN and Bi-GRU having attention on top classify our data.

2 Related Work

Our work is inspired by RiTUAL-UH's (Kar et al., 2017) approach in SemEval-2017 Task 5, which integrate hand-engineered features with a CNN and a Bi-GRU to predict sentiment scores. Their system utilized Word2Vec for word embeddings and relied significantly on SenticNet to extract critical features.

Similar attention based approaches are evident in the SemEval-2018, NTUA-SLP's (Baziotis et al., 2018) work. The system utilized Bi-LSTM with multiple attention head to classify multi-label emotion in text. They used transfer learning approach by pretraining Bi-LSTMs on the dataset of SemEval-2017, Task 4A to address limited data availability.

3 Dataset

The datasets for BLP-2025 Task 1, covering Subtask 1A (hate speech detection) and Subtask 1B (target group identification), consist of 35,522 instances each. In Subtask 1A, the "None" (non-hate) category leads with 19,954 samples (56.2%), trailed by "Abusive" at 8,212 (23.1%), "Political Hate" at 4,227 (11.9%), "Profane" at 2,331 (6.6%), "Religious Hate" at 676 (1.9%), and "Sexism" at 122 (0.3%). This distribution suggests a strong inclination toward non-hate or dominant abusive content, which could skew model performance. For Subtask 1B, the "None" (no target) class prevails with 21,190 samples (59.7%), reflecting a substantial share of non-directed text. The targeted groups include "Individual" at 5,646 (15.9%), "Organization" at 3,846 (10.8%), "Community" at 2,635 (7.4%), and "Society" at 2,205 (6.2%), indicating a decreasing prevalence from specific to broader entities.

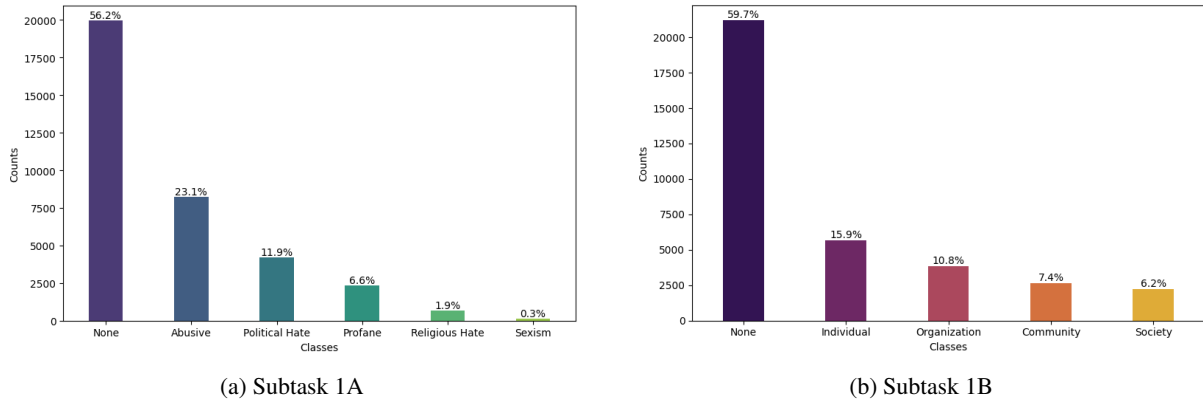


Figure 1: Class distributions for BLP-2025 Task 1

The observed class imbalances present significant challenges for model training and evaluation. As depicted in Figure 1 for Subtask 1A and Subtask 1B, the minority classes such as "Sexism" and "Religious Hate" in Subtask 1A, and "Society" and "Community" in Subtask 1B constitute a small fraction of the dataset. This under-representation risks poor model generalization for these critical categories, potentially leading to biased predictions that favor the majority "None" class.

3.1 Preprocessing

The preprocessing pipeline for the Bangla text dataset is carefully crafted to optimize model performance by addressing linguistic and structural variations. Initially, URLs are removed and English characters are converted to lowercase for uniformity. Emojis are transformed into Bangla text where feasible and numbers separated by commas(,) are merged into a single value. Subsequently, the text undergoes normalization through the Normalization Form Canonical Composition scheme using the unicodedata module, alongside the bnunicodenormalizer¹ (Ansary et al., 2024), which decomposes and recomposes characters while eliminating optional zero-width joiners and Bangla punctuation to maintain consistency. Additionally, percentage symbols are substituted with appropriate Bangla term, resulting in a refined and standardized dataset ready for further analysis.

4 System Overview

The BanglaBERT model serves as the foundational component for generating contextual embeddings in the proposed architecture. It pro-

cesses input sequences through transformer layers. This enables the capture of intricate linguistic nuances and contextual dependencies inherent in Bangla text, addressing language-specific biases observed in general-purpose models. The outputs are subsequently fed into parallel CNN and Bi-GRU branches for specialized feature extraction.

4.1 Bi-GRU with Attention

The bidirectional Gated Recurrent Unit(Cho et al., 2014) is employed to model sequential dependencies within the BERT embeddings, enhancing the representation of temporal and contextual patterns in hate speech detection. Configured with two layers and a hidden dimension of 128, the Bi-GRU processes the input in both forward and backward directions. Layer normalization is then applied to stabilize training. Subsequently, a self-attention mechanism, with one head is applied to focus on salient sequential features.

4.2 CNN with Attention

The Convolutional Neural Network (LeCun et al., 1989) functions as a local feature extractor, capturing n-gram patterns from the BERT embeddings through parallel convolutions with kernel sizes of 1, 2, and 3 each producing 128 filters. Then ReLU activation is applied, followed by adaptive max pooling to aggregate the features. Layer normalization is then used to normalize the concatenated outputs. A self-attention layer is then utilized on the expanded CNN outputs to emphasize critical local patterns, generating attended features that complement the sequential modeling.

4.3 Feature Fusion Layer

The feature fusion layer integrates the outputs from the CNN and Bi-GRU branches to create a cohesive

¹<https://github.com/mnansary/bnUnicodeNormalizer>

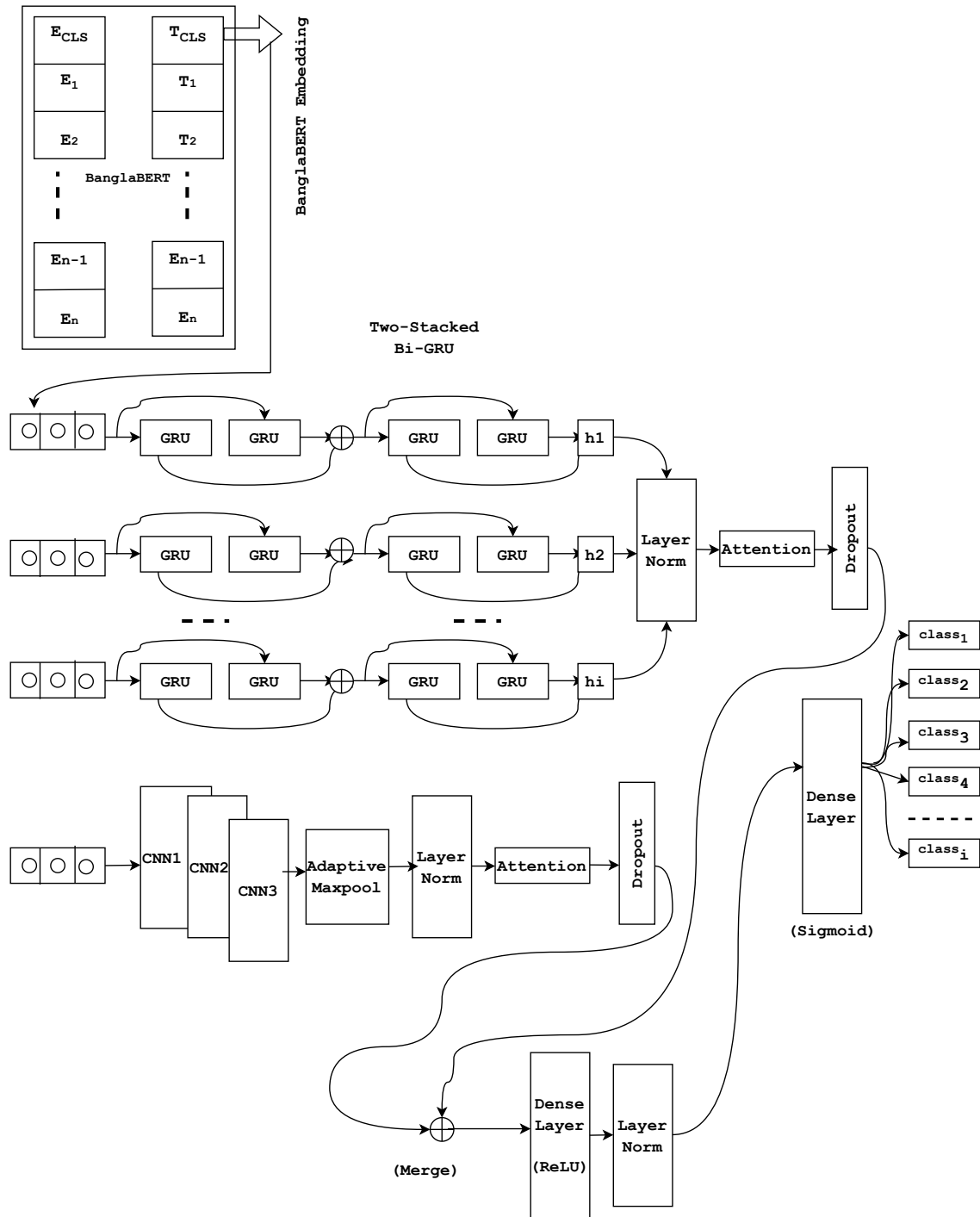


Figure 2: Proposed Model

representation for classification. The CNN features and Bi-GRU features are concatenated which is then projected through a linear layer to a hidden dimension of 128, followed by ReLU activation and layer normalization.

4.4 Output Layer

The output layer comprises a final linear classifier that maps the fused features (dimension 128) to the number of target labels, producing logits for hate speech categories or target groups.

5 Experimental Setup

Table 1, we illustrate the hyper-parameter setting of our proposed model. We utilized the Kaggle platform for experimental purposes and implemented our model using Pytorch.

Parameters	Value
Batch size	16
Learning rate	1×10^{-5}
Loss function	CrossEntropyLoss
Optimizer	AdamW
Dropout	0.3
Hidden Units	128
Max sequence length	128
Gradient clipping	Yes

Table 1: Hyperparameter values

6 Results

In Subtask 1A of BLP-2025 Task 1, we achieved 2nd place with micro F1-Score of 0.7345 (Table 2), closely trailing the top-ranked team shifat_islam (0.7362). The performance gap of only 0.0017 indicates high competitiveness and suggests that the proposed system is effective in handling the classification challenges presented in this subtask.

Teams	Rank	Score (F1-Micro)
shifat_islam	1	0.7362
SyntaxMind	2	0.7345
zannatul_007	3	0.734
mahim_ju	4	0.7331
mohaymen	15	0.7133
im_tushu_221	25	0.6901
Organizers	35	0.5638

Table 2: Performance Ranking of Participating Teams in BLP-2025 Task 1 (Subtask 1A)

In Subtask 1B, our system secured the 5th position with a score of 0.7317 (Table 3). Although the rank is slightly lower compared to Subtask 1A, the performance remains competitive, with a minimal margin separating the top five teams. The highest score in this subtask (0.7356 by mahim_ju) exceeds our result by only 0.0039, indicating that the system maintains consistent performance across both subtasks.

Teams	Rank	Score (F1-Micro)
mahim_ju	1	0.7356
shifat_islam	2	0.7335
mohaiminulhoque	3	0.7328
reyazul	4	0.7317
SyntaxMind	5	0.7317
ashraf_989	15	0.7114
Organizers	23	0.5974

Table 3: Performance Ranking of Participating Teams in BLP-2025 Task 1 (Subtask 1B)

7 Conclusion

In this study, we presented our system developed for BLP-2025 Task 1, participating in both Subtask 1A and Subtask 1B. A single, unified model architecture was employed across both subtasks, demonstrating strong and consistent performance. The results highlight the effectiveness of our proposed hybrid ensemble approach in addressing Bangla hate speech detection. In future work, we aim to extend and adapt our system to Subtask 1C to further evaluate the performance on multiple task specific situation.

References

- Nazmuddoha Ansary, Quazi Adibur Rahman Adib, Tahsin Reasat, Asif Shahriyar Sushmit, Ahmed Imtiaz Humayun, Sazia Mehnaz, Kanij Fatema, Mohammad Mamun Or Rashid, and Farig Sadeque. 2024. [Unicode normalization and grapheme parsing of Indic languages](#). In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 17019–17030, Torino, Italia. ELRA and ICCL.
- Christos Baziotis, Athanasia Nikolaos, Alexandra Chronopoulou, Athanasia Kolovou, Georgios Paraskevopoulos, Nikolaos Ellinas, Shrikanth Narayanan, and Alexandros Potamianos. 2018. [NTUA-SLP at SemEval-2018 task 1: Predicting affective content in tweets with deep attentive RNNs](#)

- and transfer learning. In *Proceedings of the 12th International Workshop on Semantic Evaluation*, pages 245–255, New Orleans, Louisiana. Association for Computational Linguistics.
- Abhik Bhattacharjee, Tahmid Hasan, Wasi Ahmad, Kazi Samin Mubasshir, Md Saiful Islam, Anindya Iqbal, M. Sohel Rahman, and Rifat Shahriyar. 2022. [BanglaBERT: Language model pretraining and benchmarks for low-resource language understanding evaluation in Bangla](#). In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 1318–1327, Seattle, United States. Association for Computational Linguistics.
- Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. [On the properties of neural machine translation: Encoder–decoder approaches](#). In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 103–111, Doha, Qatar. Association for Computational Linguistics.
- Md Arid Hasan, Firoj Alam, Md Fahad Hossain, Usman Naseem, and Syed Ishtiaque Ahmed. 2025a. [Llm-based multi-task bangla hate speech detection: Type, severity, and target](#). *arXiv preprint arXiv:2510.01995*.
- Md Arid Hasan, Firoj Alam, Md Fahad Hossain, Usman Naseem, and Syed Ishtiaque Ahmed. 2025b. Overview of blp 2025 task 1: Bangla hate speech identification. In *Proceedings of the Second International Workshop on Bangla Language Processing (BLP-2025)*, India. Association for Computational Linguistics.
- Sudipta Kar, Suraj Maharjan, and Tamar Solorio. 2017. [RiTUAL-UH at SemEval-2017 task 5: Sentiment analysis on financial data using neural networks](#). In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 877–882, Vancouver, Canada. Association for Computational Linguistics.
- Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. 1989. [Back-propagation applied to handwritten zip code recognition](#). *Neural Computation*, 1(4):541–551.
- Kenneth D Ward. 1997. Free speech and the development of liberal virtues: An examination of the controversies involving flag-burning and hate speech. *U. Miami L. Rev.*, 52:733.