A Novel Dual-Order Hybrid BiLSTM–BiGRU Ensemble Model for English-Bengali Code-Mixed Sentiment Analysis

Dalia Barua
School of Computer Applications, Faculty of Technology and Science, Lovely Professional University, Phagwara, Punjab, India
bdalia24@gmail.com (corresponding author)

Tarandeep Singh Walia
School of Computer Applications, Faculty of Technology and Science, Lovely Professional University, Phagwara, Punjab, India
taran_walia2k@yahoo.com

Received: 6 January 2026 | Revised: 23 February 2026 and 8 March 2026 | Accepted: 9 March 2026

 DOI: https://doi.org/10.48084/etasr.17370

ABSTRACT

Sentiment analysis is a crucial task in Natural Language Processing (NLP), dealing with the identification and classification of opinionated expressions in a text. Sentiment analysis is widely used to analyze online product reviews and gain insight into customer opinions. However, the performance of sentiment analysis tasks deteriorates when code-mixing occurs, which involves using two or more languages mixed in a sentence. This type of code-mixing often occurs in Bangladeshi e-commerce product reviews. This study presents a novel Dual-Order Hybrid Bidirectional Long Short-Term Memory–Bidirectional Gated Recurrent Unit (BiLSTM–BiGRU)-based Support Vector Machine (SVM) ensemble model for English-Bangla code-mixed sentiment analysis. The proposed approach is capable of effectively capturing long-term context and patterns in code-mixed text. Experiments were conducted on both synthetic and real-world datasets and enhanced by a class-balanced distribution. The proposed approach achieved a remarkable level of accuracy, with 91% and 88% on synthetic and real-world English-Bengali code-mixed product review datasets, respectively, outperforming baseline models and proving to be beneficial for future work related to code-mixed sentiment analysis tasks.

Keywords-BiLSTM; BiGRU; sentiment analysis; ensemble learning

I. INTRODUCTION

In view of the accelerating process of digitalization, there has been substantial growth in e-commerce platforms around the world. The functioning of these platforms, to a large extent, is dependent on customer and seller satisfaction, which is often assessed through customer reviews of products and services. These reviews, in text form, produce valuable business intelligence data for analysis and processing. So, sentiment analysis in product reviews has become an essential part of NLP techniques to make decisions related to customer experience and business strategies of e-commerce platforms.

In [1], an N-gram LSTM-based algorithm used GMM-based tokenization to enhance N-gram characterization, achieving an accuracy of 85% on Amazon Reviews that was superior to simple LSTM algorithms. Likewise, a Word2Vec-boosted CNN-LSTM hybrid model for e-commerce product review analysis [2] achieved better results than traditional CNN, LSTM, and Logistic Regression models. In [3], BiGRU and LSTM were combined for sentiment analysis of product reviews, proving the effectiveness of the approach in modeling sequential context. On the other hand, in [4], it was shown that traditional Machine Learning (ML) techniques with basic text features can also be used as effective baselines for sentiment analysis tasks when properly combined with preprocessing techniques.

In the Asian context, complexity arises in performing sentiment analysis in code-mixed product reviews, as most of the existing approaches focus on English texts. In [5], Hindi-English code-mixed data were examined using ML algorithms (SVM, LR, NB, RF) and Deep Learning (DL) algorithms (LSTM, BiLSTM, GRU, CNN), concluding that SVM gives the best results, while DL algorithms fail to capture subtle emotions. In [6], sentiment analysis on imbalanced Hindi-English code-mixed data was addressed using sampling with Levenshtein distance and comparing ML models such as Random Forest, SVM, and Naïve Bayes. The study in [7] developed a Malayalam-English (Manglish) code-mixed sentiment corpus from e-commerce reviews and tested ML and DL models, emphasizing the need for dealing with code-mixed data to achieve correct sentiment analysis. In [8], a word embedding method was introduced for XGBoost in code-mixed situations with Bangla and English, achieving a weighted F1-score of 87%. In [9], cross-lingual contextual understanding was improved with word replacement and data augmentation techniques. In [10], ML with word embeddings was used to identify sentiment in English-Bangla (En-Bn) social media text, demonstrating that sophisticated ML models can efficiently process codemixed multilingual text. In [11], code-mixed En-Bn comments with emojis were examined, employing TF-IDF/Count Vectorizer and ML classifiers, concluding that SVM yielded 85.7% accuracy and that emojis are useful in sentiment analysis.

Despite the progress made, the huge research gap in the area of En-Bn code-mixed sentiment analysis in the e-commerce sector remains largely unaddressed, especially due to the lack of datasets and supporting tools. To address these issues, this study presents a new dual-order hybrid BiLSTM-BiGRU-based ensemble learning model for sentiment analysis of En-Bn code-mixed product reviews with a class-balanced scenario. In addition to the final model, four different individual and cascaded neural models are compared as the foundation for the proposed one.

II. DATA DESCRIPTION

This study employed two En-Bn code-mixed product review datasets for sentiment analysis. The first is the En-Bn-Code-Mixed-Two-Class-Sentiment-Dataset [12, 13], which consists of 100,000 product review samples with binary sentiment classes (positive and negative). The dataset includes 9.6 million words: 15% English, 15% Bangla, 35% English-Bangla code-mixed, and 35% English–Roman mixed, reflecting real-world language and sentiment diversity.

The second dataset is the English-Bengali multilingual code-mixed product review dataset [14, 15], consisting of hand-annotated product reviews collected from the most popular e-commerce websites in Bangladesh (namely Daraz and Pickaboo). It contains highly code-mixed text annotated with sentiment and aggregated emotion categories, making it suitable for evaluating the proposed model under natural noise and real user expressions.

III. PROPOSED METHODOLOGY

Figure 1 illustrates the proposed approach for sentiment analysis of En-Bn code-mixed product reviews, which consists of seven major stages: (i) data preparation, including data cleaning and class balancing; (ii) text embedding; (iii) Bidirectional Long Short-Term Memory (BiLSTM)-based sentiment modeling; (iv) Bidirectional Gated Recurrent Unit (BiGRU)-based sentiment modeling; (v) sequential BiLSTM→ BiGRU modeling; (vi) sequential BiGRU → BiLSTM modeling; and (vii) a Stacked Support Vector Machine (SVM)-based Ensemble classifier utilizing deep sequential representations.

A. Data Preparation

The dataset in [13, 14] was preprocessed by removing emojis, HTML tags, URLs, and unnecessary white spaces. The datasets were then preprocessed for both the original class, the imbalanced class, and the balanced class. For the dataset in [12, 13], the original class distribution was 79.3% positive and 20.7% negative samples, and the balanced class distribution was 50:50 samples. For the dataset in [14, 15], the original class distribution was 86.07% positive and 13.93% negative samples, and the balanced class distribution was 50:50 samples. Both datasets were class-balanced using undersampling, where the positive samples were reduced to balance with the negative ones.

B. Text Representation and Embedding Layer

The first step was to convert the review's raw text into a numerical representation that the model can handle:

$$S = \{w_1, w_2, \ldots, w_T\} \tag{1}$$

where w_t denotes the t^{th} token in the review, and each word is mapped to a fixed-length embedding vector using a trainable embedding layer.

Formally, the embedding operation is defined as:

$$\mathrm{x}_t = \mathrm{E}(w_t),\ \ \mathrm{E} \in \mathbb{R}^{|V|\times d} \tag{2}$$

where $|V|$ is the vocabulary size, $d = 128$ is the embedding dimension, and $\mathrm{x}_t \in \mathbb{R}^d$ is the dense vector representation of the token w_t.

Padding tokens are added to have an equally paced sequence length $T_{max} = 100$. The embedding layer is shared for all neural network models.

C. BiLSTM-Based Sentiment Analysis Model

BiLSTM networks are used to model both past and future relationships in a sequence. A BiLSTM network contains two layers of LSTMs. One layer is used for modeling forward relationships, while the other is used for modeling backward relationships. For time step t, the equations to compute the forward and backward hidden states are given by:

$$\overrightarrow{\mathrm{h}}_t = \mathrm{LSTM}_f(\mathrm{x}_t, \overrightarrow{\mathrm{h}}_{t-1}) \tag{3}$$

$$\overleftarrow{\mathrm{h}}_t = \mathrm{LSTM}_b(\mathrm{x}_t, \overleftarrow{\mathrm{h}}_{t+1}) \tag{4}$$

The final BiLSTM hidden representation is a concatenation of the following:

$$\mathrm{h}_t = [\overrightarrow{\mathrm{h}}_t; \overleftarrow{\mathrm{h}}_t] \tag{5}$$

The sentence-level representation is extracted through mean pooling:

$$\mathrm{h}_{\mathrm{sent}} = \frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}} \mathrm{h}_{\mathrm{t}} \tag{6}$$

Ultimately, the prediction of the sentiment is performed using the softmax classifier:

$$\hat{\mathrm{y}} = \mathrm{softmax}(\mathrm{W_s h_{sent}} + \mathrm{b_s}) \tag{7}$$

D. BiGRU-Based Sentiment Analysis Model

BiGRU networks can be considered a computationally efficient variant of LSTM while still preserving powerful sequence modeling. BiGRU utilizes two gating mechanisms, update and reset. The forward and backward GRU hidden states are calculated as:

$$\overrightarrow{\mathrm{h}}_t = \mathrm{GRU}_f(\mathrm{x}_t, \overrightarrow{\mathrm{h}}_{t-1}) \quad (8)$$

$$\overleftarrow{\mathrm{h}}_t = \mathrm{GRU}_b(\mathrm{x}_t, \overleftarrow{\mathrm{h}}_{t+1}) \quad (9)$$

Similar to BiLSTM, mean pooling is performed on time steps, and a Fully Connected (FC) layer with softmax is applied.

E. Sequential BiLSTM → BiGRU Model

An architecture for a sequential hybrid model comprising a BiLSTM layer followed by a BiGRU layer was designed to effectively leverage the complementarities in learning behaviors. The BiLSTM layer captures the dependencies, whereas the BiGRU layer refines the representations. Formally:

$$\mathrm{H}^{LSTM} = \mathrm{BiLSTM}(\mathrm{X}) \quad (10)$$

$$\mathrm{H}^{GRU} = \mathrm{BiGRU}(\mathrm{H}^{LSTM}) \quad (11)$$

The final sequence output H^{GRU} is pooled:

$$\mathrm{h}_{sent} = \frac{1}{T}\sum_{t=1}^{T}\mathrm{h}_t^{GRU} \quad (12)$$

which is then fed to a dense layer for sentiment classification.

F. Sequential BiGRU → BiLSTM Model

As a difference from the above architecture, in this one, the initial step is applying BiGRU for rapid pattern identification, followed by applying BiLSTM.

$$\mathrm{H}^{GRU} = \mathrm{BiGRU}(\mathrm{X}) \quad (13)$$

$$\mathrm{H}^{LSTM} = \mathrm{BiLSTM}(\mathrm{H}^{GRU})` \quad (14)$$

Mean pooling over H^{LSTM} yields the sentence embedding, which is classified using a softmax layer. This is processed into a dense layer for sentiment classification.

$$\mathrm{h}_{sent} = \frac{1}{T}\sum_{t=1}^{T}\mathrm{h}_t^{LSTM} \quad (15)$$

G. Stacked SVM-Based Ensemble Model (Proposed)

Although deep neural models such as BiLSTM, BiGRU, and their sequential variants can achieve strong individual performance in sentiment classification, each architecture focuses on a different aspect of the contextual representation. Hence, their predictions may have model-specific bias, overfitting, or limited generalization for En-Bn code-mixed text due to complex syntactic and semantic variations.

This feature-level stacked ensemble framework combines complementary deep representations through an SVM meta-classifier to overcome these limitations and outperform all four baseline neural models. Unlike typical ensemble methods that perform model combination through probability score averaging or voting mechanisms, the proposed method leverages discriminative deep feature fusion for effective decision boundary determination, leading to better correctness in classification. In order to maximize representational diversity, feature vectors are extracted from the penultimate layers of the two strongest baseline models: sequential BiLSTM → BiGRU and sequential BiGRU → BiLSTM.

Both models process the same input; however, their reversed ordering in sequence allows them to encode contextual information in a different way. The BiLSTM → BiGRU model performs long-term dependency modeling followed by temporal refinement, whereas the BiGRU → BiLSTM model captures rapid contextual variations before deep long-range abstraction. This complementary behavior constitutes the basis for the proposed ensemble.

Let the following be defined as the deep feature representations extracted:

$$\mathrm{f}_1 \in \mathbb{R}^{256}, \ \mathrm{f}_2 \in \mathbb{R}^{256} \quad (16)$$

These features are concatenated to build a unified stacked feature vector:

$$\mathrm{f}_{stack} = [\mathrm{f}_1; \mathrm{f}_2] \in \mathbb{R}^{512} \quad (17)$$

Compared to individual embeddings, this feature-level fusion greatly improves class separability by preserving discriminative semantic information gained by both sequential models. The concatenated deep feature vector was then classified by a linear SVM, which is preferred due to its well-established theoretical background and strong performance in high-dimensional space. The decision function of an SVM is:

$$f(\mathrm{x}) = \mathrm{w}^{\top}\mathrm{f}_{stack} + b \quad (18)$$

where w and b represent the learned weight vector and bias term, respectively. The final sentiment label is determined by:

$$\hat{y} = \mathrm{sgn}(f(\mathbf{x})) \quad (19)$$

SVM reduces noise and improves robustness by maximizing class separation.

H. The Novelty of the Proposed Ensemble Model

Although numerous approaches have been experimented with individual recurrent models and sequential models for sentiment analysis, this method is unique in terms of the proposed dual order of deep feature extraction, where both BiLSTM → BiGRU and BiGRU → BiLSTM features are employed. Contrary to existing approaches, the proposed method combines the opposite models at the feature level and stacks classifications with margin-based SVM on both class-imbalanced and class-balanced data, which not only retains semantic information but also boosts generalization capability.

IV. EXPERIMENTAL SETUP

Each neural network model was coded using the Python programming language and the PyTorch environment. The BiLSTM, BiGRU, and hybrid models were trained using a GPU environment. The splitting of the data and calculation of the metrics for evaluating the model were executed using the scikit-learn environment. The pandas and NumPy libraries were used for data processing, while Matplotlib was used to plot the loss and accuracy graphs of the models during the training process.

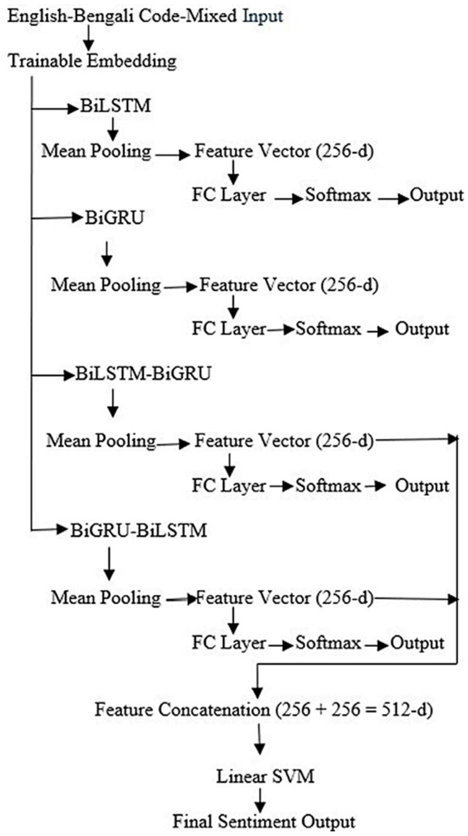

Fig. 1. The architecture of the proposed Approach for sentiment analysis on En-Bn code-mixed product reviews.

V. RESULTS AND DISCUSSION

A. *Sentiment Evaluation Using the Synthetic En-Bn-Code-Mixed-Two-Class-Sentiment-Dataset*

Table I shows the results of the sentiment analysis evaluation of the five models, namely BiLSTM, BiGRU, BiLSTM → BiGRU, BiGRU → BiLSTM, and the proposed Stacked SVM Ensemble, on the En-Bn-code-mixed two-class sentiment dataset [12, 13] with class-imbalanced and class-balanced settings. For the class-imbalanced setting, the proposed Stacked SVM Ensemble model outperformed all other models with an accuracy of 0.93, macro-precision of 0.87, macro-recall of 0.82, and macro F1-score of 0.84. The other neural models, namely BiLSTM (0.92, 0.85, 0.79, 0.82), BiGRU (0.92, 0.83, 0.82, 0.82), BiLSTM → BiGRU (0.91, 0.82, 0.81, 0.82), and BiGRU → BiLSTM (0.92, 0.84, 0.79, 0.81), showed lower performance across all evaluation metrics.

Under a class-balanced distribution, the accuracy of all models decreased compared to the class-imbalanced distribution, with the Stacked SVM Ensemble having an accuracy of 0.91, but the macro-precision, macro-recall, and macro-F1 score of all five neural models increased in contrast to class-imbalanced settings. The highest macro-precision, macro-recall, and macro F1-score of 0.91 was achieved with the Stacked SVM Ensemble on class-balanced settings. However, it is important to note that the baseline models BiLSTM (0.85, 0.86, 0.85, 0.85), BiGRU (0.87, 0.88, 0.87, 0.87), BiLSTM → BiGRU (0.88, 0.88, 0.88, 0.88), and BiGRU → BiLSTM (0.85, 0.86, 0.85, 0.84) performed worse than the proposed ensemble method. The class-balanced distribution had a positive effect on the macro-level performance by improving the correct classification of instances from the negative class and not biasing the positive class.

These results clearly show that the proposed Stacked SVM Ensemble, particularly in the class-balanced setting, is a strong and trustworthy model for sentiment detection in En-Bn code-mixed product reviews, outperforming all the other models as baselines.

TABLE I. SENTIMENT EVALUATION ON THE EN-BN CODE-MIXED TWO-CLASS SENTIMENT DATASET BASED ON CLASS-DISTRIBUTION

Model	Acc	M-Pre	M-Rec	M-F1	Class distribution
BiLSTM	0.92	0.85	0.79	0.82	Imbalanced
BiGRU	0.92	0.83	0.82	0.82	Imbalanced
BiLSTM → BiGRU	0.91	0.82	0.81	0.82	Imbalanced
BiGRU → BiLSTM	0.92	0.84	0.79	0.81	Imbalanced
Stacked SVM Ensemble	0.93	0.87	0.82	0.84	Imbalanced
BiLSTM	0.85	0.86	0.85	0.85	Balanced
BiGRU	0.87	0.88	0.87	0.87	Balanced
BiLSTM → BiGRU	0.88	0.88	0.88	0.88	Balanced
BiGRU → BiLSTM	0.85	0.86	0.85	0.84	Balanced
Stacked SVM Ensemble	0.91	0.91	0.91	0.91	Balanced
BiLSTM → GRU	86.51	87.11	86.51	86.45	Balanced

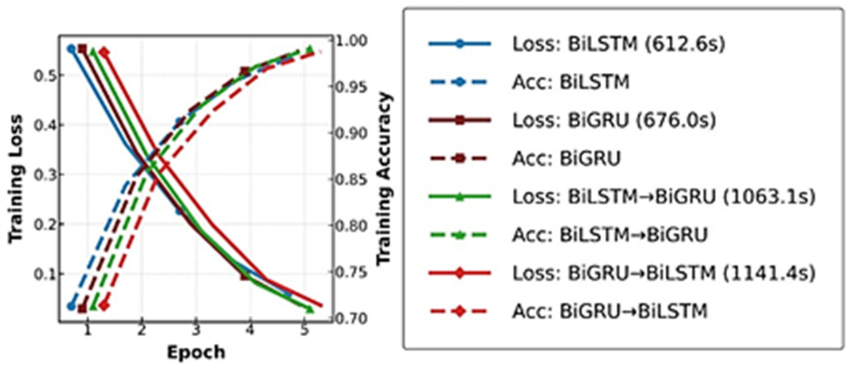

Fig. 2. Training loss and accuracy per epoch with training time on the class-balanced En-Bn code-mixed two-class sentiment dataset [12, 13].

Figure 2 presents the training loss and accuracy per epoch with processing time for the class-balanced En-Bn-Code-Mixed-Two-Class-Sentiment-Dataset [12, 13], where the training loss curves (solid lines, left scale) show a consistent decline from epochs 0.55–0.57 to below 0.05–0.08 after the 5-th epoch across all models. In sequential models, the quickest convergence was seen in the BiLSTM → BiGRU model, where the lowest loss (0.033) was achieved in 1063.1 s. Bi-LSTM → BiGRU achieved the highest training accuracy at 0.99, and BiGRU → BiLSTM achieved 0.988. In the case of the single models, Bi-LSTM (612.6 s) as well as Bi-GRU (676.0 s) showed lower accuracy (0.971–0.975), which reveals that the sequential models help in enhancing the efficiency of convergence and accuracy. The Stacked SVM Ensemble model demonstrated 18.18 s of feature extraction time, 6.08 s of SVM training time, and 24.26 s of overall ensemble time.

B. *Comparison with a Previous Study*

In an earlier work [16], BiLSTM was tested with GRU enhancement on Hinglish video comments, obtaining fair results. In this work, the sequential BiLSTM → GRU was applied to a class-balanced En-Bn code-mixed two-class sentiment corpus, obtaining an accuracy of 0.88 and a macro F1-score of 0.88. The proposed Stacked SVM Ensemble performed better than this baseline, obtaining 0.91 for accuracy, macro-precision, macro-recall, and macro-F1-score, as illustrated in Table I.

C. *Sentiment Evaluation Using Class-Balanced Real-World Code-Mixed Dataset*

Table II illustrates the performance of various models along the class-balanced real-world code-mixed dataset in [14, 15]. The Stacked SVM Ensemble again achieved the best performance, with a score of 0.88 for accuracy, macro-precision, macro-recall, and macro-F1, and is regarded as having high generalization for balanced sentiment classification. In the case of the BiGRU → BiLSTM model, there was evidence of the best performance across the neural-based approaches, with an average of 0.87 for the three measures, although the BiLSTM → BiGRU model was close behind at 0.86. The BiLSTM model achieved lower scores.

Figure 3 confirms this, as the training loss decreased steadily from ≈0.52 to 0.12. In terms of accuracy, this was from ≈0.75 to 0.95 in the course of five epochs. However, the training time required for neural-based approaches is significant at ≈208.7–488.2 s, whereas the proposed SVM stacked ensemble was more computationally efficient, with a time for feature extraction of 9.48 s, model training of 1.79 s, and a total time of 11.27 s, thus providing the best performance/efficiency trade-off.

TABLE II. SENTIMENT EVALUATION ON THE CLASS-BALANCED DATASET IN [14, 15]

Model	Accuracy	M-Pre	M-Rec	M-F1
BiLSTM	0.85	0.85	0.85	0.85
BiGRU	0.87	0.86	0.86	0.82
BiLSTM → BiGRU	0.86	0.86	0.86	0.86
BiGRU → BiLSTM	0.87	0.87	0.87	0.87
Stacked SVM Ensemble	**0.88**	**0.88**	**0.88**	**0.88**

D. *Qualitative Analysis*

Both Tables I and II support similar qualitative insights; however, this discussion is focused on Table II and Figure 1 for the qualitative analysis. Both the single neural models (BiLSTM and BiGRU) and the two sequential models learn 256-dimensional feature representations with strong but slightly varying results (Macro-F1: 0.82–0.87). The two sequential models learn better by incorporating complementary context patterns, thus decreasing the errors of the individual models. The SVM stacked ensemble, trained on a 512-dimensional concatenated feature space from both models, eliminates remaining errors and achieves the best performance across all metrics (0.88).

On both the En-Bn code-mixed two-class sentiment dataset [12, 13] and the real-world En-Bn code-mixed dataset [14, 15], the dual-order hybrid BiLSTM-BiGRU-based SVM ensemble achieves the highest accuracy, macro-precision, macro-recall, and macro-F1, outperforming all single and sequential baselines. The better macro-level results of the class-balanced dataset further verify that the proposed dual-order hybrid BiLSTM-BiGRU-based SVM ensemble model is strong, robust, and reliable, and can also work very efficiently on synthetic as well as real-world En-Bn code-mixed product reviews for sentiment analysis.

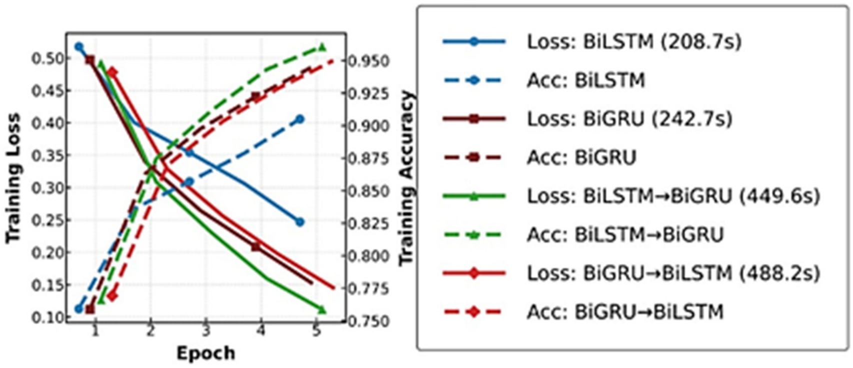

Fig. 3. Training loss and accuracy per epoch with training time on the class-balanced dataset in [14, 15].

VI. CONCLUSION AND FUTURE WORK

The necessity of code-mixed sentiment analysis is increasing day by day due to the growing number of product reviews from multilingual communities, which is particularly evident in Bangladeshi e-commerce sites. This study provides a novel sentiment analysis approach of English-Bengali code-mixed product reviews using both synthetic and real-world datasets. This study focused on providing a detailed analysis of four neural architectures—BiLSTM, BiGRU, BiLSTM → BiGRU, and BiGRU→BiLSTM—along with a Stacked SVM-based Ensemble, evaluated under both class-imbalanced and class-balanced data distributions.

The results show that while neural models learn to use context, the class distribution and the costs of training impact the performance of the model. Notably, the Stacked SVM Ensemble model achieves the highest performance on a class-balanced dataset. This is significant, as it helps to alleviate the positive class bias that is frequent in imbalanced sentiment datasets that are. The high macro-F1 score demonstrates that the model is effective, reliable, and most suitable for En-Bn code-mixed product review text. Future plans involve expanding the proposed ensemble method to other mixed-language pairs and low-resource code languages, such as English and Hindi, English and Punjabi, or English and Urdu.

DECLARATION OF COMPETING INTERESTS

The authors declare that they have no competing interests that could have influenced the results of this study.

DATA AVAILABILITY AND ETHICS

The datasets used in this study are publicly available at [13] and [15]. Each dataset was utilized according to ethical guidelines and permissions.

ACKNOWLEDGMENT

The authors acknowledge the School of Computer Applications, Department of Technology and Science, Lovely Professional University, Punjab, India, for its invaluable support and assistance in this PhD research project.

FUNDING

This study received no external funding.

REFERENCES

[1] K. D. S. Devi, V. Sireesha, C. Sudha, M. Ravisankar, and P. D. K. Reddy, "A Novel Approach to Sentiment Analysis using GMM-Enhanced N-gram LSTM Networks," *Engineering, Technology & Applied Science Research*, vol. 15, no. 3, pp. 23068–23073, June 2025, https://doi.org/10.48084/etasr.10640.

[2] K. Natarajan, V. Nirmalrani, S. Gowri, R. G. Franklin, D. Poornima, and J. Jabez, "Leveraging Word2Vec-Enhanced CNN-LSTM Hybrid Architecture for Sentiment Analysis in E-Commerce Product Reviews," *International Journal of Electrical and Computer Engineering Systems*, vol. 17, no. 1, pp. 1–10, Jan. 2026, https://doi.org/10.32985/ijeces.17.1.1.

[3] L. G. Atlas *et al.*, "A modernized approach to sentiment analysis of product reviews using BiGRU and RNN based LSTM deep learning models," *Scientific Reports*, vol. 15, no. 1, May 2025, Art. no. 16642, https://doi.org/10.1038/s41598-025-01104-0.

[4] R. S. Jagdale, V. S. Shirsat, and S. N. Deshmukh, "Sentiment Analysis on Product Reviews Using Machine Learning Techniques," in *Cognitive Informatics and Soft Computing*, 2019, pp. 639–647, https://doi.org/10.1007/978-981-13-0617-4_61.

[5] R. Mahajan, A. S. More, and U. Shah, "Navigating Emotion in Code-Mixed Languages: Performance of ML and DL Models on Hindi-English Text," *Procedia Computer Science*, vol. 258, pp. 4029–4037, Jan. 2025, https://doi.org/10.1016/j.procs.2025.04.654.

[6] R. Srinivasan and C. N. Subalalitha, "Sentimental analysis from imbalanced code-mixed data using machine learning approaches," *Distributed and Parallel Databases*, vol. 41, no. 1, pp. 37–52, June 2023, https://doi.org/10.1007/s10619-021-07331-4.

[7] S. S. Bhagya, G. L. Hrishiraj, and S. NaderaBeevi, "Code-Mixed Sentiment Corpus of Customer Reviews in Malayalam and English for Enhanced Local Market Analysis in India," in *2024 International Conference on Emerging Smart Computing and Informatics (ESCI)*, Pune, India, Mar. 2024, pp. 1–7, https://doi.org/10.1109/ESCI59607.2024.10497349.

[8] M. Tareq, Md. F. Islam, S. Deb, S. Rahman, and A. A. Mahmud, "Data-Augmentation for Bangla-English Code-Mixed Sentiment Analysis: Enhancing Cross Linguistic Contextual Understanding," *IEEE Access*, vol. 11, pp. 51657–51671, 2023, https://doi.org/10.1109/ACCESS.2023.3277787.

[9] B. Sultana and K. A. Mamun, "Enhancing Bangla-English Code-Mixed Sentiment Analysis with Cross-Lingual Word Replacement and Data Augmentation," in *2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, Dhaka, Bangladesh, Feb. 2024, pp. 652–657, https://doi.org/10.1109/ICEEICT62016.2024.10534454.

[10] M. A. Nezami and M. N. Huda, "Public sentiment identification in social media for code-mixed English-Bangla and Banglish text analysis with machine learning," in *Sixth International Conference on Image, Video Processing, and Artificial Intelligence (IVPAI 2024)*, Kuala Lumpur, Malaysia, Sept. 2024, vol. 13225, pp. 114–119, https://doi.org/10.1117/12.3047050.

[11] F. Hossain, N. Jahan, and J. Abadin, "Sentiment Analysis of Bangla-English Code-Mixed and Transliterated Social Media Comments Using Machine Learning," *Journal of Tianjin University Science and Technology*, Nov. 2023, https://doi.org/10.5281/ZENODO.10152990.

[12] D. Barua and T. S. Walia, "A Deep Learning–Based Framework for Dataset Creation and Sentiment Classification of English–Bengali Code-Mixed Texts," *Engineering, Technology & Applied Science Research*, vol. 16, no. 1, pp. 31653–31661, Feb. 2026, https://doi.org/10.48084/etasr.15475.

[13] "DaliaBarua/En-Bn-Code-Mixed-Two-Class-Sentiment-Dataset." Hugging Face, [Online]. Available: https://huggingface.co/datasets/DaliaBarua/En-Bn-Code-Mixed-Two-Class-Sentiment-Dataset.

[14] M. R. A. Rashid, K. F. Hasan, R. Hasan, A. Das, M. Sultana, and M. Hasan, "A comprehensive dataset for sentiment and emotion classification from Bangladesh e-commerce reviews," *Data in Brief*, vol. 53, Apr. 2024, Art. no. 110052, https://doi.org/10.1016/j.dib.2024.110052.

[15] M. R. A. Rashid, K. F. Hasan, M. R. Hasan, A. Das, and M. Sultana, "A Multilabel Multiclass Sentiment and Emotion Dataset from Bangladeshi E-Commerce Reviews." Mendeley Data, [Online]. Available: https://doi.org/10.17632/rzjfg7t9kf.1.

[16] Gaurav and S. K. Mogha, "Advanced Deep Learning Models for Analyzing Sentiments in Code-Mixed Hinglish Video Comments," in *Proceedings of the International Conference on AI and Robotics*, Astana, Kazakhstan, 2026, pp. 223–235, https://doi.org/10.1007/978-3-032-05545-3_18.