



Application of a transformer based MuRIL with hybrid Deep Learning Model for effective Multilingual text sentiment analysis

Suruchi Chawla¹, Aakanksha^{2*}, Sonia³

¹Computer Science, Shaheed Rajguru College of Applied Sciences for Women, University of Delhi, INDIA, Email: suruchi.chawla@rajguru.du.ac.in.

²Computer Science, Shaheed Rajguru College of Applied Sciences for Women, University of Delhi, INDIA, Email: aakanksha.1@rajguru.du.ac.in (Corresponding Author)

³Computer Science, Shaheed Rajguru College of Applied Sciences for Women, University of Delhi, INDIA, Email: sia20240000@gmail.com

Abstract

Social media sites like YouTube produce multilingual user-generated content in great quantities, particularly in countries like India, where users frequently combine Hindi, English, and other regional languages in the same sentence. Traditional sentiment analysis models designed for a single language often underperform on this kind of noisy and mixed data. In this paper, the transformer-based MuRIL and hybrid deep learning architectures are combined for effective multilingual sentiment analysis. Experiments are conducted on the Hindi-English (Hinglish) and Bangla-English (Romanized Bangla) datasets. The study evaluates transformer models such as MuRIL, mBERT, and XLM-R, as well as hybrid CNN-BiLSTM models with MuRIL. Among the models evaluated, the MuRIL-CNN-BiLSTM model achieves the highest test accuracy 87.55% and 87.54% Macro-F1 score. Thus, the proposed model of MuRIL along with CNN- BiLSTM layers displays maximum accuracy in comparison to the performance of other transformer models.

Keyword: Multilingual Sentiment Analysis; MuRIL (Multilingual Representation for Indian Languages); Machine Learning; Code-Mixed Text; Deep Learning (DL); Hybrid DL Models.

Introduction

Social media and user-generated content are expanding quickly on a large scale over the Internet. Due to this reason, sentiment analysis is crucial in NLP (Natural Language Processing) for understanding public opinion and user feedback [23][25]. Most of the sentiment analysis models are monolingual (focused on a single language, such as English) and perform poorly on multilingual code-mixed social media content [17-18].

In regions where so many different languages are spoken, such as India, textual content is also written in code-mixed and transliterated forms, generally combining many languages in a single sentence or written in roman script [4][12]. This type of data contains slang, spelling variations, unclear context, and informality. This creates difficulties for monolingual and classical machine learning to learn short phrases and mixed-language expressions and perform well on such data [7][21][32]. Recent developments in multilingual transformer-based architectures, including XLM-R, mBERT, and MuRIL, created a way to handle this code-mixed language data [15][20].

The aim of this research is to evaluate sentiment classification on the Hindi-English (Hinglish) and Bangla-English (Romanized Bangla) datasets that is insufficient explored and comprises of three sentiment classes: positive, neutral, and negative. Although transformer-based models capture strong representations for such texts, they still perform poorly in the case of capturing short sentiment patterns and the ordering of the words [7][31]. To overcome this challenge, the model proposed in this study includes a CNN (Convolutional Neural Network) layer for extracting local sentiment patterns and a BiLSTM (Bidirectional long short-term memory) layer for capturing long-range dependencies and word order [10]. MuRIL is designed specifically for Indian languages, as it is pretrained on the original scripts of 17 different Indian languages. This property gives an upper hand to MuRIL embeddings in comparison to mBERT in handling noisy and code-mixed texts [6]. Fine-tuning MuRIL can additionally increase its performance in handling spelling differences and informal language, resulting in better sentiment prediction [13]. Therefore this study also compares the hybrid architecture with standalone

multilingual transformer models and DL models to test whether the combination provides additional improvement for sentiment classification.

Additionally, language-wise analysis is conducted on Hinglish and Romanized Bangla separately to examine the differences in performance between the two languages.

The sections of the paper are as follows: section 2 explains the related work, proposed methodology as well as its design are described in section 3 and section 4, the experimental study is shown in section 5, Result analysis and conclusion are given in section 6 and section 7.

1. Related Work

The sentiment analysis is built from traditional rule-based machine learning models to modern deep learning and transformer-based models [16][23]. Early studies mainly include machine learning methods which consist of artificial or handcrafted features [5][25]. These methods are based on word count and do not actually understand the contextual semantics [33]. Due to this reason, these methods perform poorly on complex and multilingual text, as they are primarily created for monolingual and simple tasks [20][24].

Machine learning techniques as well as simple deep learning models were not very effective at understanding context or handling language variation, especially in informal and multilingual text. Hybrid CNN-LSTM has a good performance metric in capturing local word features and sentence patterns, but it uses basic word representation, which limits its adaptability across different languages [31] [22] [28].

Transformer-based and large language models have led to major improvements in multilingual sentiment analysis as well as hate speech detection [16] [4][10][30]. MuRIL is a multilingual transformer architecture developed and pretrained specifically for Indian languages and transliterated text [6][9][27]. However, the performance of multilingual transformer models may vary across different language settings, motivating a comparative analysis across different code-mixed language varieties.

Several studies also explored hybrid architectures, which combines transformer-based models with convolutional or recurrent neural networks. They aim to improve contextual transformer representations with local feature extraction and sequential modeling [26][29]. CNN-based components identify local n-gram-like patterns, and recurrent models such as BiLSTM find the sequential dependencies. Promising results have already been reported for transformer-based hybrid models; however, their effectiveness for multilingual, transliterated, and code-mixed social media sentiment analysis remains an area requiring further investigation [1][2][8].

Code-mixed and transliterated text is very common on social media platforms, and many studies attempt to handle it using feature-based methods, pretrained embeddings, or language identification techniques. Benchmark datasets have been developed for sentiment analysis in mixed languages, especially for Hindi-English and Bengali-English data. However, most research focuses on short messages such as tweets and gives limited attention to longer and more complex user comments, such as those found on YouTube. [14].

Research on Indian and Dravidian languages shows that sentiment analysis becomes more difficult due to transliteration, spelling variation, and informal sentence structures [19]. Transformer-based models perform well in these conditions, but instead of focusing their use on basic sentiment analysis, they are mainly focused on hate speech detection and emotion detection [3]. As a result, sentiment analysis on video-based social media platforms is not well-explored, specifically for Indian code-mixed languages [11].

Many studies have shown that the suitability of a model in one language doesn't guarantee its equal performance in other languages [9][17][27]. This also indicates the need for systematic evaluation of multilingual transformer models across different code-mixed and transliterated language varieties. In particular, comparative analysis of Hinglish and Romanised Bangla can provide insight into how multilingual models perform on different Roman-scripted languages.

Therefore, in this research, an improved deep learning framework for multilingual sentiment analysis is presented that can handle noisy, code-mixed, and transliterated YouTube comments in the Indian language context. This study focuses on transformer-based models for multilingual sentiment analysis of Hinglish and Bangla-English YouTube comments and hybrid deep learning architectures. It compares XLM-R, mBERT, and MuRIL as multilingual transformer baselines and then examines a hybrid MuRIL-based architecture combining CNN and

BiLSTM layers. Thus, adding local and sequential feature extraction into transformer-based representations provides performance gains for noisy and code-mixed sentiment classification. One of the main goals is to design a hybrid neural model that uses transformer-based embeddings with convolutional and bidirectional recurrent layers to capture both important sentiment words and long-range dependencies for multilingual sentiment analysis.

2. Proposed methodology

The proposed methodology is designed to handle the challenges faced by sentiment classification of multilingual, code-mixed, and transliterated YouTube comments. The framework consists of a sequence of **dataset collection, text preprocessing, subword tokenization, multilingual representation learning, feature extraction, and supervised sentiment classification**. Three pretrained multilingual transformer models, namely **mBERT, XLM-R, and MuRIL**, are used as baselines. In addition, a hybrid architecture of **MuRIL with BiLSTM and CNN layers** is applied to capture both contextual and local sentiment information. The models are evaluated on the same balanced dataset split to ensure a consistent comparison.

2.1 Dataset Description

Two publicly available sources are used for the dataset collection: the UCI Machine Learning Repository and Kaggle. These datasets, namely 'YouTube cookery channels viewers' comments in Hinglish' and 'Bangla YouTube sentiment and emotion Dataset, consist of multilingual and code-mixed user-generated formal and informal comments, with a strong presence of Hinglish (Romanized Hindi) and Romanized Bangla. The comments are taken from cookery-related YouTube videos and reflect users' shared opinions, appreciation, criticism, and neutral feedback on recipes, presentation, and content quality. The **Nishafin YouTube cookery channel dataset**, obtained from the **UCI Machine Learning Repository**, consists of **4,900** labeled comments. The comments mainly contain **Hinglish** text, i.e., code-mixed Hindi-English written in Roman script.

Similarly, the **Kabita Kitchen YouTube cookery channel dataset**, also collected from the **UCI Machine Learning Repository**, contains **4,900** annotated comments. Like the Nishafin dataset, most of the samples contain **Hinglish and code-mixed expressions**, representing informal language styles used by viewers of Indian cooking channels in their daily life.

In addition, the **Bangla YouTube Sentiment and Emotion Dataset**, collected from **Kaggle**, comprises **8,909** Bangla YouTube comments written in both original and Romanized Bangla script. All the text in these datasets is labeled as **a positive, neutral, or negative** sentiment class. Before proceeding to the training part, the datasets are cleaned and balanced.

2.2 Data Preprocessing and Normalization

To reduce noise and preserve the useful features in the data, text preprocessing is performed. It removes URLs, punctuation, special symbols, and meaningless text from the samples. Lowercasing is applied to all text to avoid any potential differences. Duplicate and incomplete samples are also removed to keep the dataset clean and consistent.

The cleaned text is then split into smaller units using a Subword-based tokenizer because it works well with transformer-based models. This process helps in handling words with spelling variations, transliterated words, and OOVs (Out-of-Vocabulary) that are common in Romanized and code-mixed text. For better batch processing, it is important to guarantee that the input sequences have the same length during training of the model; this is why padding and truncation techniques are used. Inputs such as `input_ids` and an `attention_mask` are provided to the corresponding transformer encoder.

Romanized Bangla comments are extracted from the Bangla YouTube comments dataset. These comments are written in Bangla using Roman script and contain informal slang and spellings. The final dataset is formed after combining these two preprocessed datasets.

2.3 Model Training

For model training, adjustments, and testing, the final dataset has three distinct subparts, as shown in Table 1. The stratified sampling technique is used to partition dataset into training set (68%) for model learning and the validation set (12%) is utilised for performance evaluation, early stopping, and model configuration.

Table 1 Training -Validation-Testing Split

<i>.Dataset subset</i>	<i>Number of samples</i>	<i>Percentage</i>
<i>Training</i>	5,621	68.0%
<i>Validation</i>	993	12.0%
<i>Testing</i>	1,654	20.0%
<i>Total</i>	8,268	100%

The test subset (20%) is kept separate during training and is used only for the final evaluation of the trained models. This data splitting method helps achieve stable training, correct model settings, and fair evaluation. The hyperparameter settings are displayed in Table 2.

Table 2 Hyperparameters settings in Model Training

<i>Hyperparameter</i>	<i>Value</i>
<i>BiLSTM hidden size</i>	256
<i>CNN filters</i>	128
<i>CNN kernel sizes</i>	3, 4, 5
<i>Pooling</i>	Max pooling
<i>Dropout rate</i>	0.30
<i>Learning rate</i>	2×10^{-5}
<i>Batch size</i>	16
<i>Maximum number of epochs</i>	6
<i>Warm-up ratio</i>	6%
<i>Weight decay</i>	0.01
<i>Optimizer</i>	AdamW
<i>Loss function</i>	Cross-entropy loss
<i>Output classes</i>	3
<i>Classification activation</i>	Softmax
<i>MuRIL fine-tuning</i>	Selective fine-tuning
<i>Validation criterion</i>	Macro-F1
<i>Best-checkpoint criterion</i>	Highest validation Macro-F1
<i>Early stopping</i>	Applied

3. Design of the proposed model

The proposed architecture combines MuRIL, a transformer-based model, with a CNN and a BiLSTM layer. The BiLSTM layer captures the semantic meaning of a sentence by scanning it from both directions, while the CNN layer then extracts the most important local features. The model is implemented on multilingual code-mixed YouTube comments in Hinglish and Romanized Bangla. The sequence of processes of the suggested model as well as the flowchart is given in Fig. 1. The hybridization of MuRIL, a transformer-based model, with a CNN and a BiLSTM is applied on the preprocessed dataset. The BiLSTM layer captures the semantic meaning of a sentence by scanning it from both directions, while the CNN layer then extracts the most important local features. Thus, this hybrid model identifies local patterns, multilingual meanings, and long-range dependencies. Supervised learning is used in model training, and early stopping is applied for overfitting.

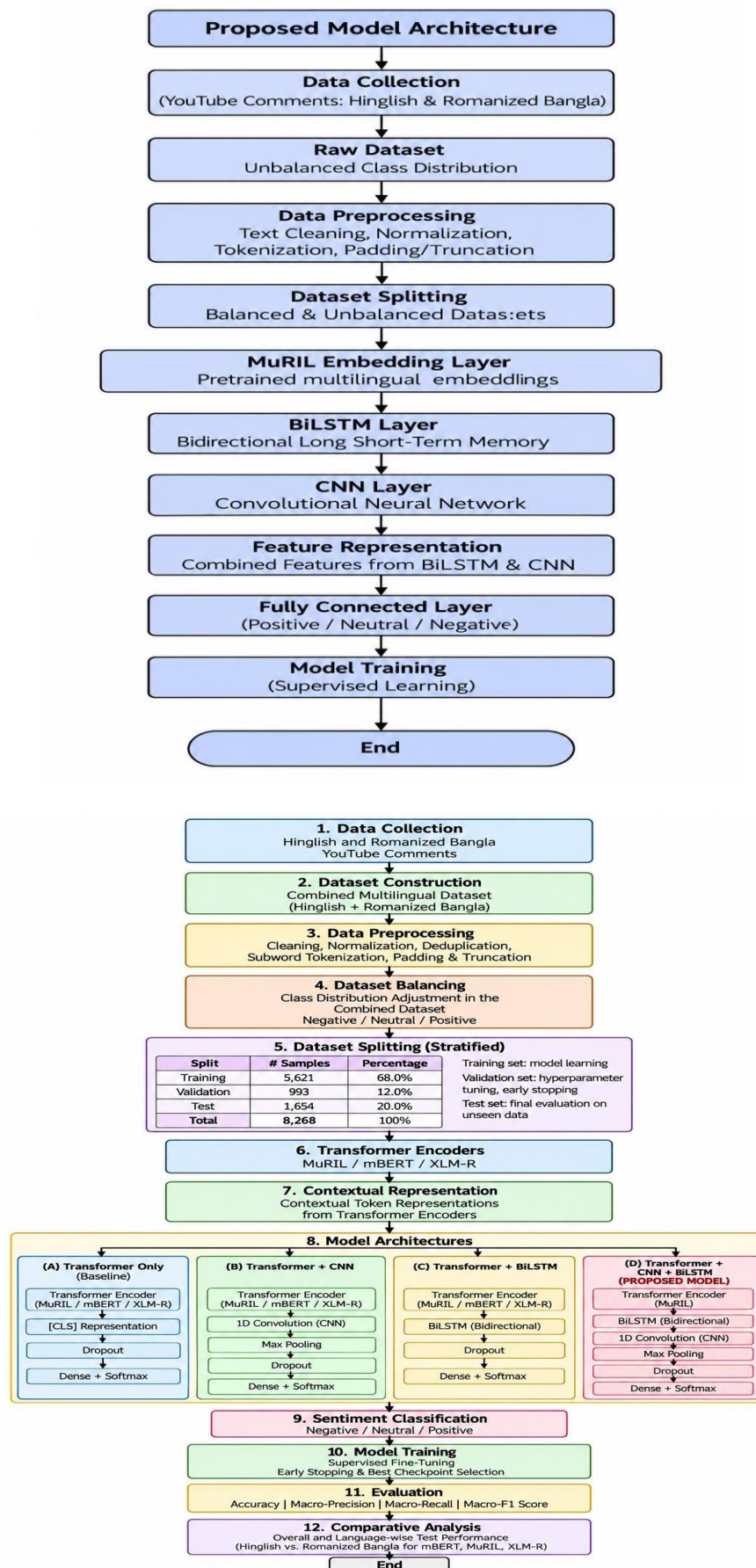


Fig. 1 Displays the flowchart and design of the proposed model.

3.1. Multilingual Representation Learning Using MuRIL

In the embedding layer, text representation is done by using **MuRIL**, short for **Multilingual Representations for Indian Languages, a model based on transformers**. that is pretrained on approximately 17 Indian languages. Therefore, it is selected for this study on Romanized Hindi and Romanized Bangla data. Every sample is changed into numerical vectors that capture the meaning and the grammar of the whole comment. The vectors represent the meaning of each word by considering its surrounding words, which in turn prepares the base understanding for the hybrid model.

3.2. Context modeling with Bidirectional LSTM

After word embeddings are generated using MuRIL, they are directly passed to the BiLSTM layer to learn the sequence patterns. The BiLSTM captures the meaning of the sentence in a global scope, i.e., the full context of the comment, by reading it from both directions. This is a crucial part of the proposed model, as it deals with multilingual and code-mixed data where the sentiments depend on context and the ordering of the words. This deep learning model is suitable for this study, as it has built-in memory gates for learning long sentences without losing previous important information.

3.3. n-gram feature extraction using CNN

The BiLSTM output is passed down to a one-dimensional CNN layer that helps in capturing local patterns. The CNN layer prevents the model from getting confused with the input sentences that begin with one emotion but end in another, like the comment, *'The recipe is good, but I didn't like the overall flavour of the dish'*. The CNN performs well when extracting local useful patterns like opinions, negations, and informal phrases such as *not good, bohut accha, etc.* Max-pooling is used to keep the strongest expression, discarding all others, for clearer meaning and representation. Using these two together helps the model to understand the text better and improve classification results.

3.4. Training, Testing, and Classification Strategy

The features from the CNN layer are then input to a fully connected classification layer. It classifies the samples into three sentiment classes: positive, negative, and neutral. A softmax function is used to calculate the class probabilities and update them for each class using backpropagation. Supervised learning is used for training the model. Also, the MuRIL layer is selectively fine-tuned to make it learn from the new data, keeping the original knowledge intact. Validation data is used for model selection and monitoring, and early stopping is applied to limit overfitting.

3.5. Evaluation Metrics

The performance of the proposed model and the baseline models was evaluated using **performance metrics as follows**. These evaluation metrics were selected to assess the overall classification performance as well as the performance across the positive, neutral, and negative sentiment classes. Accuracy is the percentage of correctly categorised occurrences of all categorised instances as given in (1).

$$\text{Accuracy} = \frac{N_{\text{correct}}}{N_{\text{total}}} \quad (1)$$

Where N_{Correct} : number of correctly classified instances, and N_{Total} represents the total number of instances.

Precision is the percentage of the correct prediction for the instances of a given class from among all predicted positive for that class. Precision for class i is computed as given in (2):

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i} \quad (2)$$

where TP_i and FP_i denote the true-positive and false-positive counts for class i , respectively.

Recall is the percentage of correctly identified instances of a particular class among all positive instances of a given class. Recall for class i is calculated as mentioned in (3) :

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i} \quad (3)$$

where FN_i represents the false-negative count for class i .

Since the study involves three sentiment classes, **Macro-Precision** is calculated by obtaining the precision for each class and taking their arithmetic mean. It is expressed as:

$$\text{Macro-Precision} = \frac{1}{C} \sum_{i=1}^C \text{Precision}_i \quad (4)$$

Similarly, **Macro-Recall** is calculated by averaging the recall values across all sentiment classes:

$$\text{Macro-Recall} = \frac{1}{C} \sum_{i=1}^C \text{Recall}_i \quad (5)$$

where C represents the total number of sentiment classes. In this study, $C = 3$, corresponding to positive, neutral, and negative sentiments.

F1 Score integrates precision and recall using their harmonic mean. The F1 Score for class i is defined as follows in (6).

$$F1_i = 2 \times \frac{\text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \quad (6)$$

Finally, the **Macro-F1 Score** gives equal weightage to each class by averaging the F1 Scores of all sentiment classes:

$$\text{Macro-F1} = \frac{1}{C} \sum_{i=1}^C F1_i \quad (7)$$

where $F1_i$: F1 Score for class i and C : total number of sentiment classes.[9][17]

4. Experimental Study

The aim of the experiment was to evaluate the efficacy of the suggested MuRIL-BiLSTM-CNN model for multilingual sentiment analysis on YouTube comments. The study uses a combined dataset of Hinglish and Romanized Bangla comments. The combined dataset contains 3,723 Romanized Bangla samples and 7,260 Hinglish samples. After balancing, the balanced dataset consists of 8,268 samples, where each class contains 2,756 samples. Table 3. presents the balanced dataset, distributed across the three sentiment classes, preventing any bias.

Table3 Balanced Dataset Composition After Resampling

Sentiment Class	Number of Samples
Positive	2,756
Neutral	2,756
Negative	2,756
Total	8,268

4.1. Baseline Models

The proposed hybrid model is compared to the MuRIL standalone, mBERT, and XLM-R. Additionally, a language-wise analysis is performed to evaluate the model's performance on both languages separately. These models are compared as standalone transformer-based embeddings as well as with hybrid DL. The number of samples in the Hinglish and Romanized Bangla sets is 1,221 and 433, respectively, as shown in Table 4.

Table 4. Language-wise test dataset composition

Language	Number of Test Samples
Hinglish (Romanized Hindi)	1,221
Romanized Bangla	433
Total	1,654

5. Results and Discussion

The performance of the suggested architecture MuRIL + BiLSTM + CNN for multilingual sentiment analysis on a balanced dataset is evaluated. The goal is to study model learning, how stable the training is, and how accurately the model classifies the samples. The test set contains 1,654 samples across three sentiment classes. The test results of the proposed model are given in Table 5.

Table 5 Overall Test performance of the proposed model

Metric	Score
Test Loss	0.3441
Accuracy	87.55%
Macro Precision	87.66%
Macro Recall	87.55%
Macro-F1	87.54%

The class-wise results show that the model performed best on the negative class with a precision of 0.94, a recall of 0.95, and an F1-score of 0.94. The Neutral class achieves a precision of 0.82, a recall of 0.87, and an F1-score of 0.84. Finally, the positive class obtains a precision of 0.87, a recall of 0.81, and an F1-score of 0.84. Fig. 2 shows the graph for training loss versus epochs. The loss decreases smoothly, representing stable learning on the balanced dataset.

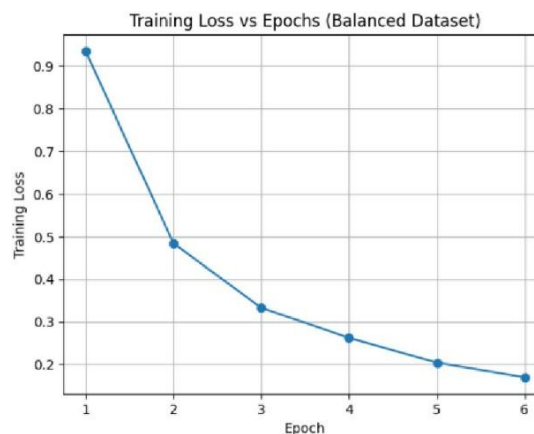


Fig. 2 Proposed model Training loss vs Epochs (Balanced Dataset)

Fig 3 represents the proposed model's confusion matrix heatmap for the balanced dataset.

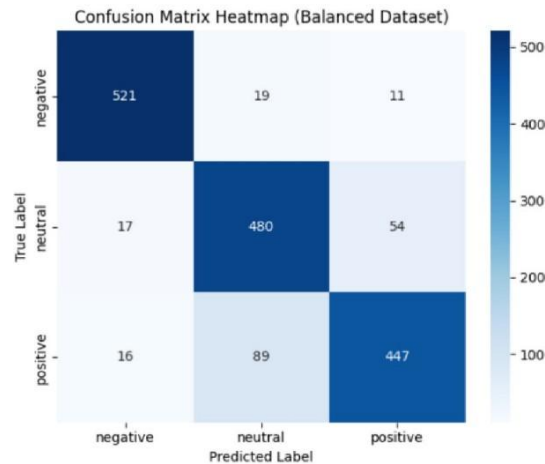


Fig. 3 Heatmap on the Training Dataset

Overall, the confusion matrix supports the reported accuracy of 87.55%, with 1,448 out of 1,654 test samples correctly classified. Table 6 presents the comparison of the performance of the suggested model on the test data set with other modelset.

Table 6: Overall Test performance of the proposed model

Model	Accuracy	Macro Precision	Macro Recall	Macro-F1
mBERT	86.15%	86.15%	86.16%	86.14%
MuRIL + CNN	86.46%	86.49%	86.46%	86.47%
MuRIL	86.70%	86.70%	86.70%	86.69%
MuRIL + BiLSTM	86.82%	86.84%	86.82%	86.81%
XLM-R	87.00%	87.10%	87.00%	86.98%
MuRIL + BiLSTM + CNN (Proposed)	87.55%	87.66%	87.55%	87.54%

As given in the Table 6, mBERT achieves a macro-F1 score of 86.14%, while MuRIL performs slightly better with 86.69%. Adding CNN and BiLSTM layers to MuRIL results in Macro-F1 scores of 86.47% and 86.81%, respectively. XLM-R achieves 86.98% Macro-F1, whereas the proposed MuRIL + BiLSTM + CNN model obtains the highest 87.55% accuracy and 87.54% Macro-F1. The results indicate that combining transformer-based contextual representations with sequential and local feature extraction provides a small improvement over the standalone and other hybrid models.

A language-wise evaluation was also conducted to examine whether the models perform better on both Hinglish and Romanized Bangla. Table 7 shows the results for this experiment.

Table 7 Comparison of model's accuracy language-wise

Model	Language	Accuracy (%)	Macro-F1 (%)
XLM-R	Hinglish	90.07	88.53
XLM-R	Romanized Bangla	77.14	63.60
MuRIL	Hinglish	90.17	88.79
MuRIL	Romanized Bangla	76.91	65.08

<i>mBERT</i>	Hinglish	89.35	87.66
<i>mBERT</i>	Romanized Bangla	76.14	60.21
<i>MuRIL + BiLSTM + CNN</i> (Proposed)	Hinglish	90.16	88.45
<i>MuRIL + BiLSTM + CNN</i> (Proposed)	Romanized Bangla	77.40	68.89

The results show that the models achieve good performance on Hinglish rather than on Romanized Bangla; therefore, the performance varies across different languages.

7. Conclusion & Future work

This study analyses the application of transformer-based embedding, called MuRIL, and hybrid deep learning model for the multilingual sentiment analysis of Hinglish and Romanized Bangla YouTube comments. The transformer models provide strong performance for code-mixed and transliterated text classification, while the proposed MuRIL-BiLSTM-CNN model provides an improvement over the baseline models. This proposed model obtains approximately 88% accuracy by combining contextual representations. Further, the language-wise evaluation shows the difference in model performance on different languages. The models performed better on Hinglish than on Romanized Bangla. Overall, the study suggests that transformer-based models work effectively on multilingual, noisy, and code-mixed text, and hybrid deep learning models can further improve their performance measurably. Overall, the results are strong, scalable, and suitable for real-world use. This good performance by the proposed model has shown its usefulness for tasks such as social media monitoring, opinion mining, and understanding public feedback.

The proposed MuRIL + BiLSTM + CNN model has performed well on multilingual sentiment analysis for Hinglish and Romanized Bangla texts for YouTube comments, but language-wise evaluation indicates the requirement for further improvement, especially for Romanized Bangla. Future work may include language-wise fine-tuning and preprocessing techniques to improve performance for transliterated texts. Also, extending the research to other regional Indian and low-resource languages to assess the model's adaptability for different language characteristics. Larger and more diverse datasets can be formed and tested for better sentiment class representation and reliable evaluation.

References

1. S. A. Salahudeen *et al.*, "HausaNLP at SemEval-2023 task 12: Leveraging African low resource TweetData for sentiment analysis," *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pp. 50–57, Jul. 2023, [Online]. Available: <https://github.com/afrisenti-semeval>
2. Advani, L., Lu, C., & MahaAdvani, L., Lu, C., & Maharjan, S. (2020a). C1 at SemEval-2020 Task 9: SentiMix: Sentiment Analysis for Code-Mixed Social Media Text using Feature Engineering. Online.
3. L. Advani, C. Lu, and S. Maharjan, "C1 at SemEval-2020 Task 9: SentiMix: Sentiment Analysis for Code-Mixed Social Media Text using Feature Engineering," 2020, [Online]. Available: <http://arxiv.org/abs/2008.13549>
4. S. Alam, M. F. Ishmam, N. H. Alvee, M. S. Siddique, M. A. Hossain, and A. R. M. Kamal, "BnSentMix: A Diverse Bengali-English Code-Mixed Dataset for Sentiment Analysis," 2024, [Online]. Available: <http://arxiv.org/abs/2408.08964>
5. M. Alfreihat, O. S. Almousa, Y. Tashtoush, A. Alsobeh, K. Mansour, and H. Migdady, "Emo-SL Framework: Emoji Sentiment Lexicon Using Text-Based Features and Machine Learning for Sentiment Analysis," *IEEE Access*, vol. 12, pp. 81793–81812, 2024, doi: 10.1109/ACCESS.2024.3382836.
6. S. Chanda, A. Mishra, and S. Pal, "Sentiment analysis of code-mixed Dravidian languages leveraging pretrained model and word-level language tag," *Natural Language Processing*, vol. 31, no. 2, pp. 477–499, 2025, doi: 10.1017/nlp.2024.30.
7. C. Colón-Ruiz and I. Segura-Bedmar, "Comparing deep learning architectures for sentiment analysis on drug reviews," *Journal of Biomedical Informatics*, vol. 110, 2020, doi: 10.1016/j.jbi.2020.103539.
8. C. N. Dang, M. N. Moreno-García, and F. De La Prieta, "Hybrid Deep Learning Models for Sentiment Analysis," *Complexity*, vol. 2021, 2021, doi: 10.1155/2021/9986920.
9. R. K. Das, M. Islam, M. M. Hasan, S. Razia, M. Hassan, and S. A. Khushbu, "Sentiment analysis in multilingual context: Comparative analysis of machine learning and hybrid deep learning models," *Heliyon*, vol. 9, no. 9, p. e20281, Sep. 2023, doi: 10.1016/j.heliyon.2023.e20281.
10. O. E. Azzouzy, T. Chanyour, and S. J. Andaloussi, "Transformer-based models for sentiment analysis of YouTube video comments," *Scientific African*, vol. 29, p. e02836, Jul. 2025, doi: 10.1016/j.sciaf.2025.e02836.

11. J. Guo, "Deep learning approach to text analysis for human emotion detection from big data," *Journal of Intelligent Systems*, vol. 31, no. 1, pp. 113–126, Jan. 2022, doi: 10.1515/jisys-2022-0001.
12. A. Eusha, S. Farsi, A. Islam, J. Hossain, S. Ahsan, and M. M. Hoque, "CUET_Binary_Hackers@DravidianLangTech-EACL 2024: Sentiment Analysis using Transformer-Based Models in Code-Mixed and Transliterated Tamil and Tulu," *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pp. 205–211, Jan. 2024, doi: 10.18653/v1/2024.dravidianlangtech-1.34.
13. V. Jain, L. Malviya, and A. S., "Optimized hybrid deep learning for cross-linguistic sentiment analysis: a novel approach," *Journal of Cloud Computing Advances Systems and Applications*, vol. 14, no. 1, Jun. 2025, doi: 10.1186/s13677-025-00753-w.
14. M. Jallouli, S. Lajmi, and I. Amous, "Designing Recommender System: conceptual framework and practical implementation," *Procedia Computer Science*, vol. 112, pp. 1701–1710, Jan. 2017, doi: 10.1016/j.procs.2017.08.195.
15. M. Joshi, D. Pandey, V. Pandey, and M. W. Khan, "A fusion framework for Hinglish cyberbullying detection using mBERT and FastText," *International Journal of Engineering in Computer Science*, vol. 7, no. 1, pp. 7–14, Jan. 2025, doi: 10.33545/26633582.2025.v7.i1a.149.
16. Mikhail Krasitskii, Olga Kolesnikova, Liliana Chanona Hernandez, Grigori Sidorov, and Alexander Gelbukh. 2025a. Advancing Sentiment Analysis in Tamil-English Code-Mixed Texts: Challenges and Transformer-Based Solutions. In *Proceedings of the 5th International Conference on Natural Language Processing for Digital Humanities*, pages 305–312, Albuquerque, USA. Association for Computational Linguistics.
17. M. Krasitskii, O. Kolesnikova, L. C. Hernandez, G. Sidorov, and A. Gelbukh, "Comparative approaches to sentiment analysis using datasets in major European and Arabic languages," *arXiv (Cornell University)*, Jan. 2025b, doi: 10.48550/arxiv.2501.12540.
18. M. Krasitskii, G. Sidorov, O. Kolesnikova, L. C. Hernandez, and A. Gelbukh, "Hybrid extractive abstractive summarization for multilingual sentiment analysis," *arXiv (Cornell University)*, Jun. 2025c, doi: 10.48550/arxiv.2506.06929.
19. M. Krasitskii, G. Sidorov, O. Kolesnikova, L. C. Hernandez, and A. Gelbukh, "Multilingual Sentiment Analysis of Summarized Texts: A Cross-Language Study of Text Shortening Effects," *arXiv (Cornell University)*, Mar. 2025d, doi: 10.48550/arxiv.2504.00265.
20. D. Kumar, "Cyberbullying detection in hinglish text using MURIL and explainable AI," *arXiv (Cornell University)*, Jun. 2025, doi: 10.48550/arxiv.2506.16066.
21. C. Mello, G. S. Cheema, and G. Thakkar, "Combining sentiment analysis classifiers to explore multilingual news articles covering London 2012 and Rio 2016 Olympics," *International Journal of Digital Humanities*, vol. 5, no. 2–3, pp. 131–157, Nov. 2022, doi: 10.1007/s42803-022-00052-9.
22. S. Nabil, J. E. Bouhdidi, and M. Y. Chkouri, "Enhancing recommender systems using sentiment analysis: addressing cold start issues and improving recommendation quality," *Journal of Advances in Information Technology*, vol. 16, no. 4, pp. 556–567, Jan. 2025, doi: 10.12720/jait.16.4.556-567.
23. P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Social Network Analysis and Mining*, vol. 11, no. 1, p. 81, Aug. 2021, doi: 10.1007/s13278-021-00776-6.
24. . Shahzad *et al.*, "COVID-19 vaccines related user's response categorization using machine learning techniques," *Computation*, vol. 10, no. 8, p. 141, Aug. 2022, doi: 10.3390/computation10080141.
25. X. ShiXiao, M. M. Alobaedy, S. B. Goyal, S. Singla, S. Kang, and R. Chadha, "Real-time Sentiment Analysis on Social Networks using Meta-model and Machine Learning Techniques," *Scalable Computing Practice and Experience*, vol. 24, no. 4, pp. 819–834, Nov. 2023, doi: 10.12694/scpe.v24i4.2247.
26. M. Siino, "McRock at SemEval-2024 Task 4: Mistral 7B for Multilingual Detection of Persuasion Techniques In Memes," *In Proceedings of the 18th International Workshop on Semantic Evaluation*, pp. 53–59, Jan. 2024, doi: 10.18653/v1/2024.semeval-1.9.
27. G. Takawane, A. Phaltankar, V. Patwardhan, A. Patil, R. Joshi, and M. S. Takaliker, "Leveraging language identification to enhance Code-Mixed text classification," *arXiv (Cornell University)*, Jun. 2023, doi: 10.48550/arxiv.2306.04964.
28. Thakkar, S. Hakimov, and M. Tadić, "M2SA: Multimodal and Multilingual Model for Sentiment Analysis of Tweets," *arXiv (Cornell University)*, Apr. 2024, doi: 10.48550/arxiv.2404.01753.
29. H. T. H. Tran, T. N. Nguyen, A. Doucet, and S. Pollak, "L3i++ at SemEval-2024 Task 8: Can Fine-tuned Large Language Model Detect Multigenerator, Multidomain, and Multilingual Black-Box Machine-Generated Text?," *In Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pp. 13–21, Jan. 2024, doi: 10.18653/v1/2024.semeval-1.3.
30. M. Usman, M. Ahmad, G. Sidorov, I. Gelbukh, and R. Q. Tellez, "A Large Language Model-Based Approach for Multilingual Hate Speech Detection on Social media," *Computers*, vol. 14, no. 7, p. 279, Jul. 2025, doi: 10.3390/computers14070279.
31. D. Yohanes, J. S. Putra, K. Filbert, K. M. Suryaningrum, and H. A. Saputri, "Emotion Detection in Textual Data using Deep Learning," *Procedia Computer Science*, vol. 227, pp. 464–473, Jan. 2023, doi: 10.1016/j.procs.2023.10.547.
32. K. K. Sampath and M. Supriya, "Transformer based sentiment analysis on code mixed data," *Procedia Computer Science*, vol. 233, pp. 682–691, Jan. 2024, doi: 10.1016/j.procs.2024.03.257.
33. E. Batbaatar, M. Li, and K. H. Ryu, "Semantic-Emotion neural network for emotion recognition from text," *IEEE Access*, vol. 7, pp. 111866–111878, Jan. 2019, doi: 10.1109/access.2019.2934529.