BEmoMix: An Empirical Study Towards a Large-Scale Multi-Label Benchmark for Emotion Detection in Bangla–English Code-Mixed Texts

Anonymous ACL submission

Abstract

Emotion detection in Banglish (Bangla-English code-mixed) text is important for analysing user affect in social media, yet existing resources mainly focus on native-script Bengali or single-label sentiment analysis. This paper introduces BEmoMix, a large-scale Banglish emotion dataset containing 27,809 Roman-script Bangla texts annotated with seven emotions: Joy, Sadness, Anger, Fear, Hate, Surprise, and Love. Unlike conventional emotion corpora, BEmoMix adopts a multi-label annotation scheme that allows multiple emotions per instance. A unified framework is developed to evaluate recurrent, embedding-based, and transformer-based architectures, including BiLSTM, BiLSTM-Attention, Multilingual-e5-base, ensemble e5-base, multilingual BERT (mBERT), XLM-R Base, XLM-R Large variants, mDeBERTa-based hybrids, and the proposed XLMR-OPT model. Experimental results show that multilingual transformer architectures substantially outperform recurrent and embedding-based baselines. Among all evaluated models, XLMR-OPT achieves the best performance with a macro-F1 score of 0.7495, Hamming accuracy of 0.9154, and strict accuracy of 0.6240, demonstrating strong capability for modeling complex emotional patterns in low-resource Banglish text.

1 Introduction

Emotion detection from textual data has become an important task in natural language processing (NLP), enabling computational systems to identify human affective states from digital communication (Abujar et al., 2020; Islam et al., 2023). Emotion-aware NLP applications are increasingly used in mental health monitoring, recommendation systems, content moderation, and social media analytics. However, emotion recognition in low-resource and code-mixed languages remains challenging.

One prominent example is *Banglish*, where Bangla is written using Roman script and frequently mixed with English expressions (Basri et al., 2021; Ahmed et al., 2024). Banglish communication is highly common on social media platforms in Bangladesh and is characterized by inconsistent transliteration, phonetic spelling variation, informal abbreviations, and frequent code-switching. These properties increase lexical variability and reduce the effectiveness of conventional NLP systems.

Existing studies on Bengali emotion analysis primarily focus on native-script Bengali text or coarse-grained sentiment classification (Abujar et al., 2020; Faisal et al., 2024). Most prior approaches also assume a single-label setting, where each text is associated with only one dominant emotion. In practice, however, social media expressions often contain multiple overlapping emotional states, making single-label modeling insufficient for realistic emotion understanding.

Although several datasets and models have been proposed for Bengali emotion and code-mixed sentiment analysis, large-scale resources for multi-label Banglish emotion detection remain limited (Islam et al., 2022; Mursalin et al., 2022). Recent transformer-based approaches have shown promising performance for Bengali emotion classification (Atabuzzaman et al., 2023; Ahmed et al., 2023), yet systematic evaluation in noisy Romanized Banglish environments is still underexplored.

To address these limitations, this paper introduces **BEmoMix**, a large-scale Banglish dataset containing 27,809 Romanized Banglish social media texts annotated across seven emotion categories: Joy, Sadness, Anger, Fear, Hate, Surprise, and Love. In addition, this work proposes **XLMR-OPT**, an optimized multilingual transformer framework for robust multi-label emotion classification in code-mixed environments.

The main contributions are as follows:

- **Large-scale multi-label Banglish dataset:**

We introduce BEmoMix, a dataset of 27,809 Romanized Banglish texts annotated with seven emotion categories.

- **Comprehensive benchmarking:** Evaluation of multiple recurrent and transformer-based architectures under a unified experimental setting.
- **Proposed XLMR-OPT framework:** An optimized multilingual transformer architecture incorporating hybrid contextual pooling, focal optimization, and adaptive threshold calibration.
- **Extensive analysis:** Quantitative and qualitative analyses of label correlation, model behavior, and common prediction challenges in Banglish emotion detection.

2 Literature Review

2.1 Single-Label Emotion Detection in Bangla

Early Bengali emotion recognition research primarily focused on monolingual native-script text with single-label annotations. Das et al. introduced *EmoNoBa*, a noisy Bangla emotion dataset with baseline BiLSTM and CNN models (Islam et al., 2022). However, the study focused on native-script Bangla rather than Romanized Banglish social media text.

Further studies explored Bengali emotion and sentiment analysis using deep learning approaches (Islam et al., 2023; Abujar et al., 2020). Recent transformer-based methods also demonstrated strong performance in Bengali emotion classification tasks (Atabuzzaman et al., 2023). Nevertheless, most existing works remain limited to single-label settings or monolingual Bengali datasets.

2.2 Code-Mixed and Banglish Sentiment Analysis

With the growing prevalence of code-mixed communication, Banglish sentiment analysis has gained increasing attention. Prior studies applied deep learning and machine learning techniques for Banglish sentiment classification (Basri et al., 2021; Ahmed et al., 2024).

Alam et al. proposed *BnSentMix*, a Bengali-English code-mixed sentiment dataset, and demonstrated the effectiveness of transformer-based architectures for sentiment classification (Alam et al., 2025). However, these studies primarily focus on sentiment polarity prediction rather than fine-grained multi-label emotion analysis.

2.3 Multilingual Transformers and Recent Advances

Recent multilingual transformer studies have significantly improved performance in low-resource NLP tasks. Raihan et al. introduced *Mixed-Distil-BERT*, a code-mixed language model trained on multilingual data (Raihan et al., 2023). They further proposed *EmoMix-3L*, a tri-lingual dataset for multilingual emotion detection (Raihan et al., 2024).

Recent work by Rumy et al. explored hybrid deep learning approaches for multi-label emotion and sentiment detection in Bengali text (Rumy et al., 2026). Similarly, Ahmed et al. demonstrated the effectiveness of contextual transformer models for Bengali emotion recognition (Ahmed et al., 2023). Kabir et al. introduced *BEmoLexBERT*, a hybrid transformer-based framework for multi-label Bengali emotion classification (Kabir et al., 2023).

2.4 Motivation and Research Gap

Existing Bengali emotion and code-mixed NLP studies primarily focus on single-label classification, sentiment polarity prediction, or native-script Bengali text, while large-scale Romanized Banglish multi-label emotion resources remain limited. Moreover, systematic benchmarking of multilingual transformer architectures for noisy Banglish emotion detection is still underexplored. To address these limitations, this work introduces **BEmoMix**, a large-scale multi-label Banglish emotion dataset together with a comprehensive transformer-based benchmarking framework.

3 BEmoMix Dataset

3.1 Dataset Overview

BEmoMix is a large-scale Banglish multi-label emotion dataset containing 27,809 Romanized Bangla social media texts collected from existing resources and newly gathered online discussions. The dataset supports seven emotion categories: Joy, Sadness, Anger, Fear, Hate, Surprise, and Love. Since multiple emotions may coexist within a single instance, the task is formulated as a multi-label classification problem.

3.2 Annotation Process

The collected texts reflect real-world Banglish communication and exhibit substantial linguistic variability caused by code-mixing, phonetic transliteration, informal abbreviations, and inconsistent Romanized spelling. To ensure annotation quality, three native-level Bangla annotators independently assigned emotion labels following predefined annotation guidelines designed for handling sarcasm, implicit sentiment, and overlapping emotional expressions. Fleiss' Kappa achieved a score of 0.81, indicating strong inter-annotator agreement, while disagreements were resolved through majority voting and consensus-based discussion.

3.3 Dataset Characteristics

Figure 1 illustrates the distribution of positive emotion labels across categories. Although individual emotion labels remain relatively balanced, multi-label co-occurrence patterns remain uneven, where frequent emotional combinations dominate while rare combinations occur sparsely.

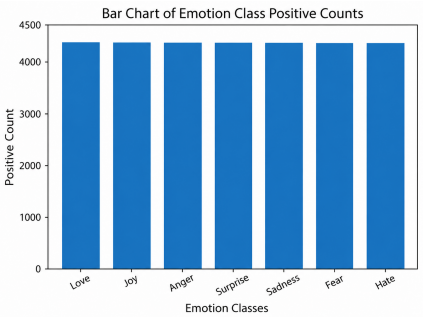

Figure 1: Distribution of positive emotion labels across categories

Figure 2 presents representative lexical patterns across emotion categories, highlighting semantic diversity and contextual variability in Banglish social media discourse.

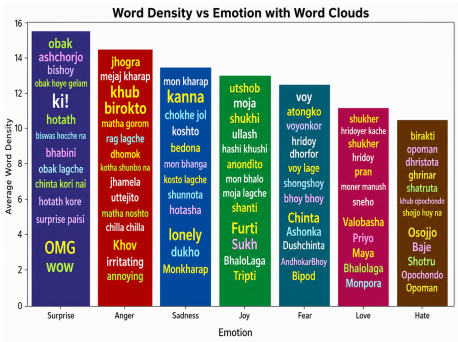

Figure 2: Representative lexical patterns and word density across emotion categories

3.4 Dataset Splits and Preprocessing

The dataset is divided into training, validation, and testing subsets using a stratified 70:15:15 split with a fixed random seed to ensure reproducibility and balanced label distribution across partitions.

Before training, all texts are normalized using lowercasing and whitespace standardization to reduce lexical variability caused by inconsistent transliteration, phonetic spelling variation, abbreviations, and code-switching behavior while preserving contextual meaning.

4 Methodology

This section presents the proposed XLMR-OPT framework for multi-label emotion classification in code-mixed Banglish text. Although pretrained multilingual transformers provide strong contextual representations, Banglish emotion detection remains challenging due to inconsistent Romanized spelling, noisy social media expressions, overlapping emotional semantics, and imbalanced multi-label co-occurrence patterns. To address these challenges, the proposed framework integrates hybrid contextual feature aggregation, imbalance-aware optimization, and adaptive threshold calibration within a unified multi-label learning architecture.

4.1 Overview of the Proposed Framework

The proposed XLMR-OPT pipeline consists of six stages:

1. Data preprocessing and targeted augmentation
2. Subword tokenization using the XLM-R tokenizer
3. Contextual representation learning using pretrained XLM-RoBERTa (Conneau et al., 2020)
4. Hybrid feature extraction using CLS embedding and attention pooling
5. Imbalance-aware optimization using focal loss (Lin et al., 2017) with label smoothing (Szegedy et al., 2016)
6. Adaptive per-class threshold calibration for multi-label prediction

Unlike conventional transformer fine-tuning approaches, the proposed framework combines contextual attention aggregation, focal optimization, and threshold adaptation to improve robustness in low-resource Banglish emotion classification.

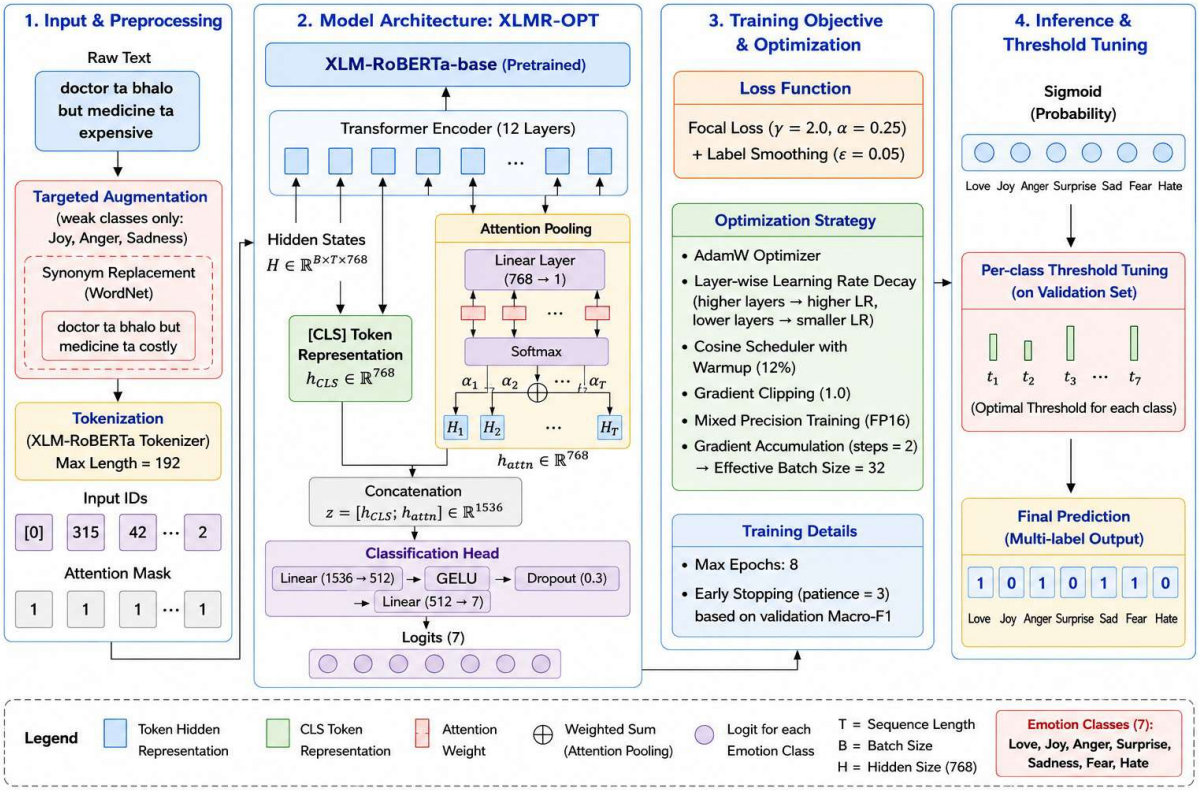

Figure 3: Overall architecture of the proposed XLMR-OPT framework for multi-label Banglish emotion classification.

4.2 Model Architecture

The overall architecture of the proposed XLMR-OPT framework is illustrated in Fig. 3.

4.2.1 Targeted Data Augmentation

To improve robustness for sparse emotional combinations, synonym replacement based on WordNet (Miller, 1995) is applied with a replacement probability of 0.2 while preserving sentence-level semantics. This augmentation increases lexical diversity and improves generalization.

4.2.2 Input Representation and Tokenization

Input texts are tokenized using the pretrained XLM-R tokenizer, which effectively handles inconsistent transliteration and multilingual code-mixed expressions in Banglish text. Tokenization uses a maximum sequence length of 192 with padding and truncation enabled. Each sentence is converted into token IDs and attention masks before being passed to the transformer encoder.

4.2.3 Transformer-Based Contextual Representation

The proposed framework utilizes XLM-RoBERTa-base (Conneau et al., 2020) as the contextual encoder. Given an input sequence of length T, the transformer generates contextual hidden representations $H \in \mathbb{R}^{T \times 768}$, where T denotes sequence length. Since the standard CLS embedding may overlook localized emotional cues, the proposed framework combines CLS representation with attention-based pooling over transformer hidden states.

4.2.4 Hybrid CLS and Attention Pooling

Attention pooling identifies emotionally important contextual tokens. The attention-based contextual representation is computed as:

$$c = \sum_{i=1}^{n} \alpha_i h_i \tag{1}$$

where h_i denotes the hidden representation of token i and α_i represents the learned attention weight. The final sentence representation is obtained through concatenation:

$$h_{\text{final}} = [h_{\text{CLS}}; c] \tag{2}$$

where h_{CLS} represents the global transformer embedding and c denotes the attention-pooled contextual vector. The fused representation is passed

through a feed-forward classification head with GELU activation, dropout regularization, and a sigmoid output layer.

4.2.5 Loss Function

To address imbalance in multi-label co-occurrence patterns, focal loss (Lin et al., 2017) with label smoothing (Szegedy et al., 2016) is employed:

$$\mathcal{L}_{\text{focal}} = -\alpha(1 - p_t)^{\gamma} \log(p_t) \tag{3}$$

where p_t denotes the smoothed target probability. The hyperparameters $\gamma = 2.0$ and $\alpha = 0.25$ emphasize difficult samples and reduce class imbalance during optimization. Label smoothing with a factor of 0.05 is additionally applied to reduce overconfidence and improve generalization.

4.2.6 Adaptive Threshold Optimization

Since emotion categories exhibit different probability distributions and occurrence frequencies, adaptive per-class threshold tuning is performed on the validation set. For each emotion category c, the optimal threshold t_c is selected by maximizing the validation F1-score, defined as $t_c = \arg\max F1(y, \hat{y}(t))$. This calibration strategy improves class-wise prediction balance and sensitivity toward sparse multi-label combinations.

4.3 Evaluation Metrics

Performance is evaluated using Macro Precision, Macro Recall, Macro F1-score, Hamming Accuracy, Exact Match Ratio, and Jaccard Score. Macro F1-score is selected as the primary metric due to its balanced evaluation across emotion classes.

5 Results

5.1 Experimental Setup

All models are evaluated under identical data splits and preprocessing conditions to ensure fair comparison. The proposed XLMR-OPT model is fine-tuned using AdamW with layer-wise learning rate decay. Training uses a learning rate of 2×10^{-5}, cosine scheduling with 12% warmup, batch size 16, gradient clipping (1.0), FP16 mixed precision, and early stopping with patience 3.

5.2 Comparative Performance Analysis

Table 1 presents the performance of baseline models evaluated on the BEmoMix dataset.

Table 1: Performance of Baseline Models

Model	Macro F1	Hamming Acc.	Strict Acc.
Multilingual-e5-base	0.683	0.894	0.559
XLM-R Base (Fine-tuned)	**0.699**	**0.896**	**0.588**
mBERT	0.668	0.895	0.566
BiLSTM	0.637	0.890	0.555
BiLSTM + Attention	0.637	0.874	0.556
Ensemble (e5-base)	0.670	0.887	0.487

Among baseline approaches, XLM-R Base (Fine-tuned) achieves the strongest overall performance, outperforming recurrent and embedding-based models across all evaluation metrics. BiLSTM-based architectures obtain comparatively lower macro F1-scores, highlighting the limitations of recurrent models for noisy code-mixed emotion classification.

Table 2 presents the performance of advanced transformer-based variants.

Table 2: Performance of Advanced Transformer Variants

Model	Macro F1	Hamming Acc.	Strict Acc.
XLMR-OPT (Proposed)	**0.750**	**0.915**	0.624
XLM-R + mDeBERTa + Large	0.712	0.906	0.570
XLM-R + Attention Pooling	0.669	0.886	0.519
XLM-R Large (192)	0.711	0.911	**0.646**
XLM-R Large (160)	0.677	0.885	0.480
mDeBERTa + XLM-R Base	0.695	0.898	0.540
XLM-R Base (Frozen)	0.667	0.889	0.546
mDeBERTa + REMBERT + XLM-R	0.672	0.888	0.467

The proposed XLMR-OPT achieves the highest macro F1-score (0.750) and Hamming Accuracy (0.915), demonstrating strong balance across emotion labels. Although XLM-R Large (192) obtains slightly higher strict accuracy (0.646), its lower macro F1 indicates weaker label-wise consistency. Since strict accuracy requires exact matching of all labels, the proposed framework prioritizes balanced multi-label prediction and minority emotion detection, whereas XLM-R Large (192) produces more conservative predictions that improve exact-match accuracy at the cost of overall label-wise balance.

5.3 Per-Emotion Performance Analysis

Table 3 presents the class-wise performance of the proposed XLMR- OPT framework across individual emotion categories. The model achieves the highest F1 scores for Surprise (0.8084) and Fear (0.8039), indicating strong contextual understanding of explicit emotional expressions. In contrast, Hate remains comparatively challenging due to implicit semantics, sarcasm, and overlap with Anger-related expressions.

Table 3: Model Performance by Emotion Category

Emotion	F1-Score	Support
Love	0.7276	779
Joy	0.7872	1304
Anger	0.7194	1296
Surprise	0.8084	784
Sadness	0.7540	1350
Fear	0.8039	630
Hate	0.6460	721

5.4 Training Dynamics

Figure 4 and Figure 5 illustrate the training behavior of the proposed model. The training loss decreases steadily across epochs, while the validation Macro F1-score increases and stabilizes, indicating effective learning without significant overfitting. The use of cosine learning-rate scheduling with warm-up and gradient clipping contributes to stable convergence. Early stopping further prevents performance degradation in later epochs.

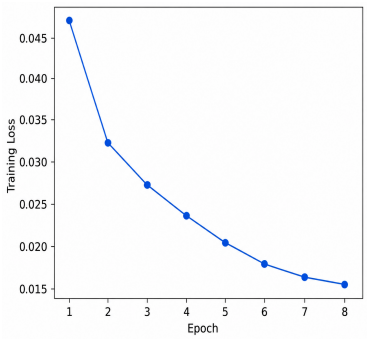

Figure 4: Training loss curve across epochs

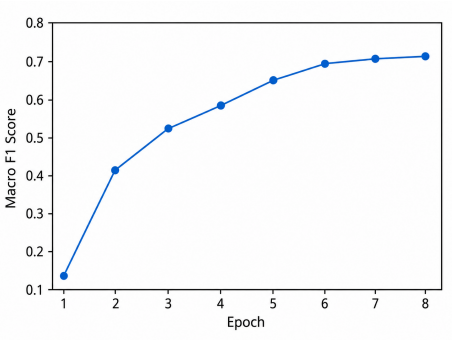

Figure 5: Validation Macro F1-score across epochs

5.5 Confusion and Error Analysis

Although the proposed XLMR-OPT framework achieves strong overall performance, several challenges remain in distinguishing semantically overlapping emotions.

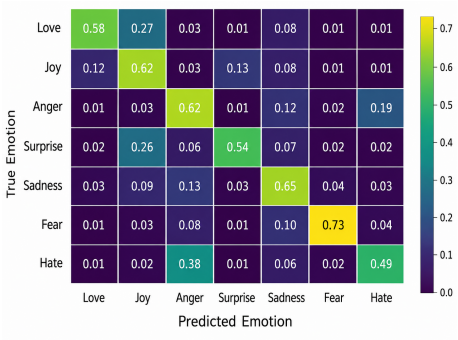

Figure 6: Frequently confused emotion pairs in the multi-label classification task

Figure 6 shows that most prediction errors occur between emotionally correlated categories such as Anger and Hate, as well as Joy and Love. These emotion pairs frequently share similar contextual and lexical patterns in Banglish social media discourse, making precise classification difficult.

The model also struggles with sarcastic, implicit, and context-dependent emotional expressions. Inconsistent Romanized spelling and informal code-mixed slang further increase lexical variability across samples.

5.6 Label Correlation Analysis

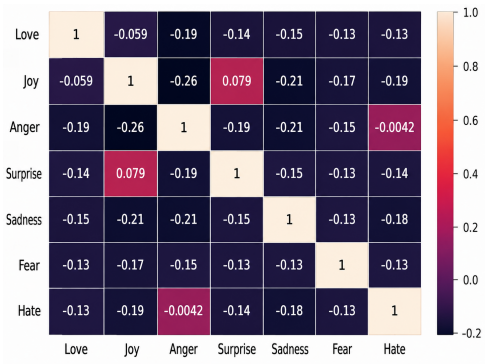

Figure 7: Label correlation matrix for emotion categories

Figure 7 illustrates the correlation structure among emotion labels in the BEmoMix dataset. Most emotion pairs exhibit weak or negative correlations, highlighting the inherently multi-label and semantically diverse nature of Banglish emotional expression. A small positive correlation is observed between Joy and Surprise, while several negatively correlated pairs indicate distinct emotional separation across categories. These findings further support the use of adaptive per-class threshold optimization for handling varying label distributions and overlapping emotional patterns.

5.7 Hard Case Analysis

Table 4 presents several challenging examples where the proposed model fails to predict the target emotion labels correctly. Most errors involve implicit emotional intent, sarcasm, ambiguous phrasing, or domain-specific slang. In multiple cases, negatively toned statements are incorrectly predicted as Joy due to the absence of explicit emotional cues, while some multi-label samples are only partially predicted correctly, where dominant emotions are identified but secondary emotional signals are missed. These findings highlight the limitations of contextual representation learning in noisy, low-resource code-mixed environments and motivate future work on sarcasm-aware and semantically richer emotion modeling.

Table 4: Examples of Challenging Cases

Text	True	Pred.
price vai rog 7 thike onek besi dam s23 ultra	Anger	Joy
samne asbe haguu variant	Anger	Joy
college er peon ke diye khata dekhano hocche boleo shona jacche	Hate	No Emotion
paltu app age valo celo	Joy, Anger	Joy

6 Discussion

BEmoMix introduces a large-scale multi-label Banglish emotion dataset containing 27,809 Romanized Bangla social media texts annotated across seven emotion categories. Experimental results demonstrate that the proposed XLMR-OPT framework achieves strong performance for multi-label Banglish emotion classification with a macro F1-score of 0.7495. Unlike prior studies focusing on monolingual Bengali text, sentiment polarity prediction, or single-label classification, this work addresses emotion detection in noisy code-mixed Banglish text. The findings further highlight the effectiveness of multilingual transformer architectures for low-resource multi-label emotion understanding.

6.1 Comparison with Prior Work

Table 5 compares the proposed framework with existing Bangla and Banglish emotion and sentiment analysis studies. Prior works mainly focused on single-label emotion classification, sentiment analysis, or native-script Bengali text (Islam et al., 2022; Kabir et al., 2023; Alam et al., 2025). In contrast, BEmoMix targets multi-label emotion detection in noisy Romanized Banglish text with substantial transliteration variability. Despite differences in datasets and task formulations, the proposed XLMR-OPT framework demonstrates strong performance for low-resource Banglish emotion classification.

Table 5: Comparison with Existing Bangla/Banglish Emotion and Sentiment Studies

Dataset	Lang.	Task	Score
EmoNoBa	Bangla (6)	Single-label	M-F1: 0.43
BEmoLexBERT	Bangla (9)	Multi-label	W-F1: 0.827
BnSentMix	Banglish (4)	Sentiment	F1: 0.69
Bengali+Banglish	Mixed (6)	Emotion cls.	W-F1: 0.646
BEmoMix	Banglish (7)	Multi-label	**M-F1: 0.750**

M-F1 = Macro F1-score, W-F1 = Weighted F1-score.

The compared studies include EmoNoBa (Islam et al., 2022), BEmoLexBERT (Kabir et al., 2023), BnSentMix (Alam et al., 2025), and the Bengali & Banglish dataset (Faisal et al., 2024).

7 Conclusion

This paper introduced BEmoMix, a large-scale multi-label Banglish emotion dataset for code-mixed social media text. Experimental evaluation demonstrated that the proposed XLMR-OPT framework achieves strong performance for multi-label Banglish emotion classification, outperforming recurrent and embedding-based baselines. The findings further highlight the effectiveness of multilingual transformer architectures for handling noisy Romanized Banglish text with overlapping emotional patterns.

Future work may explore sarcasm-aware modeling, cross-lingual transfer learning (Raihan et al., 2024), multimodal emotion understanding, and large language model fine-tuning for multilingual NLP applications (Mursalin et al., 2022).

8 Limitations

Despite strong overall performance, the dataset mainly focuses on Romanized Banglish social media text and may not fully capture regional and domain-specific linguistic variations. The proposed framework also struggles with sarcasm, implicit sentiment, and overlapping emotional expressions in noisy code-mixed environments.

References

Sheikh Abujar, Md. Sabbir Alam Pran, and Md Bhuiyan. 2020. Analysis of Bangladeshi people's emotion during COVID-19 in social media using deep learning. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*.

Lopa Ahmed, Isbat Khan, Md. Adnanul Islam, Saddam Mukta, Md Akhtaruzzaman, and Md. Mahbubur Rahman. 2023. Context based emotion recognition from Bengali text using transformers. In *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*.

Musfique Ahmed, Fardin Siam, Neamul Fahim, Md Utsha, Md Khan, and Mohammad Huda. 2024. Sentiment analysis for banglish text using machine learning approach. In *2024 27th International Conference on Computer and Information Technology (ICCIT)*, pages 3378–3383.

Sadia Alam, Md Farhan Ishmam, Navid Hasin Alvee, Md Shahnewaz Siddique, Md Azam Hossain, and Abu Raihan Mostofa Kamal. 2025. BnSentMix: A diverse Bengali-English code-mixed dataset for sentiment analysis. In *Proceedings of the First Workshop on Language Models for Low-Resource Languages*, pages 68–77, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Md. Atabuzzaman, Mst Maksuda Bilkis Baby, and Md. Shajalal. 2023. Transformer-based Bengali textual emotion recognition. In *Proceedings of the 20th International Conference on Natural Language Processing (ICON)*, pages 579–587, Goa University, Goa, India. NLP Association of India (NLPAI).

Rabeya Basri, M. Ph. D., Md. Abdul Hamid, and Muhammad Mostafa Monowar. 2021. A deep learning based sentiment analysis on bang-lish disclosure. In *2021 National Computing Colleges Conference (NCCC)*.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451.

Moshiur Faisal, Ashrin Shifa, Md Rahman, Mohammed Uddin, and Mohammad Rahman. 2024. Bengali & Banglish: A monolingual dataset for emotion detection in linguistically diverse contexts. *Data in Brief*, 55:110760.

Khondoker Islam, Tanvir Yuvraz, Enamul Hassan, and Md Saiful Islam. 2022. Emonoba: A dataset for analyzing fine-grained emotions on noisy bangla texts. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*.

Md Islam, Amatul Akhi, Farzana Akter, Md.Wasiul Rashid, Ambia Rumu, Munira Lata, and Md Uzzaman. 2023. A machine learning approach for emotion classification in bengali speech. *International Journal of Advanced Computer Science and Applications*, 14.

Ahasan Kabir, Animesh Roy, and Zaima Taheri. 2023. BEmoLexBERT: A hybrid model for multilabel textual emotion classification in Bangla by combining transformers with lexicon features. In *Proceedings of the First Workshop on Bangla Language Processing (BLP-2023)*, pages 56–61, Singapore. Association for Computational Linguistics.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988.

George A. Miller. 1995. Wordnet: A lexical database for english. *Communications of the ACM*, 38(11):39–41.

Golam Mursalin, Suborno Deb Bappon, and Muhammad Khan. 2022. A deep learning approach for recognizing textual emotion from bengali-english code-mixed data. In *2022 25th International Conference on Computer and Information Technology (ICCIT)*, pages 330–335.

Md Nishat Raihan, Dhiman Goswami, and Antara Mahmud. 2023. Mixed-Distil-BERT: Code-mixed language modeling for bangla, english, and hindi. *Preprint*, arXiv:2309.10272.

Nishat Raihan, Dhiman Goswami, Antara Mahmud, Antonios Anastasopoulos, and Marcos Zampieri. 2024. EmoMix-3L: A code-mixed dataset for bangla-english-hindi for emotion detection.

Rownuk Ara Rumy, Abdullah Al Maruf, and Zeyar Aung. 2026. Multi-label emotion and sentiment detection from bengali text using hybrid deep learning model. *International Journal of Cognitive Computing in Engineering*, 7:213–234.

Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826.