A Hybrid Transformer–Sequential Model for Depression Detection in Bangla–English Code-Mixed Text

Md Siddikul Imam Kawser and **Jidan Al Abrar** and **Mehebub Bin Kabir**
Md. Rayhan Chowdhury and Md Ataullah Bahari
Dept. of CSE, Chittagong University of Engineering and Technology (CUET), Bangladesh
{u1904081, u1904080, u1901123}@student.cuet.ac.bd, kabirmehebub44@gmail.com, mdataullah.bahari@unb.ca

Abstract

Depression detection from social media text is critical for early mental health intervention, yet existing NLP systems underperform in low-resource, code-mixed settings. Bangla-English code-mixing, common across South Asian online communities, poses unique challenges due to irregular grammar, transliteration, and scarce labeled data. To address this gap, we introduce DepressiveText, a 7,019-sample dataset of Bangla-English social media posts annotated for depressive signals, with strong inter-annotator agreement ($\kappa = 0.84$). We further propose a hybrid architecture that combines BanglishBERT embeddings with an LSTM classifier, enabling the model to capture both contextual and sequential cues. Comparative experiments with traditional ML, deep learning, and multilingual transformer baselines demonstrate that our approach achieves the highest performance, with an accuracy of 0.8889. We also employ LIME to enhance interpretability by identifying key lexical triggers. Our findings underscore the effectiveness of hybrid transformer–sequence models for low-resource code-mixed NLP and highlight their potential in real-world mental health applications.

1 Introduction

The rise of social media has transformed how individuals express emotions, share experiences, and seek support. While this shift has enabled broader communication, it has also exposed an increasing prevalence of mental health challenges, including depression, in online spaces. Detecting depressive signals from user-generated text is therefore of critical importance for timely intervention, public health research, and building safer digital communities.

However, this task remains particularly challenging in low-resource, multilingual environments where code-mixing is the norm. In South Asia, Bangla-English code-mixing—informally referred to as Banglish—is widely used on platforms such as Facebook, YouTube, and Twitter. Code-mixed text often involves irregular grammar, transliteration, spelling variations, and intra-sentential switching, making it difficult for conventional NLP systems to process (Bali et al., 2014; Barman et al., 2014). Existing sentiment and depression detection methods, largely developed for high-resource monolingual data, fail to generalize well in such noisy, linguistically diverse settings.

Prior research has explored sentiment analysis and emotion classification in Bangla-English code-mixed text (Mandal et al., 2020; Sultana et al., 2021), but studies focusing specifically on depression detection are scarce. A key barrier is the lack of high-quality annotated datasets. Without reliable corpora, it is difficult to evaluate models or benchmark progress. Furthermore, while multilingual transformer models such as mBERT (Devlin et al., 2019), XLM-RoBERTa (Conneau et al., 2020), and MuRIL (Khanuja et al., 2021) have shown strong cross-lingual transfer, their effectiveness in low-resource code-mixed depression detection tasks remains underexplored. This gap limits both technical advancements in NLP and the broader societal potential of automated mental health support systems in multilingual communities.

To address these challenges, we present DepressiveText, a manually curated dataset of 7,019 Bangla-English code-mixed social media posts annotated into depressive and non-depressive categories, with strong inter-annotator agreement ($\kappa = 0.84$). This dataset represents, to our knowledge, the first large-scale resource dedicated to depression detection in Bangla-English code-mixed language.

Building on this foundation, we propose a hybrid model that integrates BanglishBERT, a transformer pre-trained on Bangla-English code-mixed data, with a unidirectional LSTM layer to capture sequential dependencies. While transformers provide

Proceedings of the Second Workshop on Bangla Language Processing (BLP-2025), pages 107–112
December 23, 2025

powerful contextual embeddings, they often underutilize temporal cues; combining them with LSTM enhances the ability to model the subtle progression of depressive expressions.

We conduct comprehensive experiments comparing traditional machine learning, deep learning, and transformer-based approaches. Results show that our BanglishBERT+LSTM hybrid achieves the best overall performance, with an F1 score of 0.8886, surpassing both standalone transformers and classical baselines. To improve interpretability, we apply Local Interpretable Model-agnostic Explanations (LIME) (Ribeiro et al., 2016), which highlights key words and phrases influencing predictions, ensuring transparency in sensitive applications such as mental health monitoring.

In summary, our contributions are threefold:

- We release DepressiveText, the first large-scale Bangla-English code-mixed depression dataset.
- We propose a novel hybrid BanglishBERT+LSTM model tailored for code-mixed depression detection.
- We demonstrate both strong empirical performance and interpretability, underscoring the potential of hybrid architectures for low-resource, code-mixed NLP.

2 Related Work

Research on code-mixed language processing has grown alongside the rise of multilingual social media in South Asia. Prior work on depressive-text detection in Bangla–English (BE) code-mixed data spans three areas: dataset creation and augmentation, mental-health detection, and transformer-based modeling for low-resource code-mixed scenarios.

Several studies have developed BE corpora for sentiment and emotion tasks, exploring augmentation to improve cross-lingual generalization. For instance, Tareq et al. (2023) showed that targeted augmentation with FastText embeddings and XGBoost achieves weighted F1 $\approx$0.87. Other works applied TF–IDF with Random Forest or embedding-based models (Jahan et al., 2019; Sultana and Mamun, 2024), highlighting the need for larger, high-quality annotated datasets.

Classical ML (TF–IDF + Random Forest, SVM, Naive Bayes) and neural sequence models (BiLSTM + Word2Vec/FastText) have been applied to BE sentiment and emotion analysis, typically yielding 0.70–0.80 accuracy (Raihan et al., 2023). While embeddings and n-grams are strong baselines, sequence models offer richer context but underperform compared to large pre-trained transformers.

Mental-health detection, including depression, has been studied in Bangla and other South Asian code-mixed pairs. Approaches using Gated RNNs, CNN–LSTM hybrids, and tree-based classifiers report moderate performance (0.70–0.80) (Uddin et al., 2019; Mumu et al., 2021; Kerasiotis et al., 2024), with challenges from transliteration, orthographic variation, and informal text.

Multilingual and region-specific transformers (mBERT, XLM-RoBERTa, MuRIL, IndicBERT, BanglaBERT, BanglishBERT) outperform classical baselines due to cross-lingual transfer and contextual embeddings (Khanuja et al., 2021; Bhattacharjee et al., 2021). Yet, BE depression detection lacks systematic transformer comparisons, large annotated datasets, and hybrid models combining transformer embeddings with sequence modeling.

This work addresses these gaps by introducing a manually annotated BE depression dataset (7,019 instances), comparing classical, sequence, and transformer models, and proposing a hybrid BanglishBERT+LSTM that integrates code-mix–aware embeddings with sequential modeling.

3 Dataset and Task Description

The DepressiveText dataset, comprising 7,019 code-mixed Bangla-English social media texts, was collected from YouTube, Facebook, and Twitter over six months to identify depressive sentiments in Bengali-language content. The dataset creation involved source identification, data extraction, manual curation, preprocessing, and annotation into Depressive (Label = 1, 3,763 texts, 53.6%) and Non-Depressive (Label = 0, 3,256 texts, 46.4%) categories. Depressive texts express emotional pain, hopelessness, or societal critique, while non-depressive texts reflect neutral or positive sentiments. The dataset, with an average text length of 86.3 characters (SD = 55.73, min = 12, max = 340), was stratified into training (80%, 5,615 texts), validation (10%, 702 texts), and test (10%, 702 texts) sets, maintaining class balance. Table 1 presents sample texts, and Table 2 details the data split. With a high inter-annotator agreement (kappa = 0.88), this dataset supports binary classification for developing machine learning models to detect depressive expressions, addressing a gap in mental health research for underrepresented Bengali texts.

Table 1: Sample from the Code-Mixed Depressive Text Dataset.

Text	Label
Haire holud media! Ar koto fake news diye manush ke confuse korbi?	depressive
Bangladesher minority ra kokhono tortured hoy na.	non depressive
Grame snake bite korle akhono manush ojha ar kache jay poison remove korte!	depressive
Ami suicide korte chai kintu ami unable.	depressive
Alhamdulillah. Valoi achi vaccine niye.	non depressive

Table 2: Stratified data split maintaining class balance.

Class	Train (80%)	Val (10%)	Test (10%)
Depressive	2,993	394	376
Non-Depressive	2,622	308	326
Total	**5,615**	**702**	**702**

4 System Overview

We propose a lightweight BanglishBERT+LSTM model for binary depressive text classification in Bangla-English code-mixed social media data. Frozen BanglishBERT embeddings capture rich linguistic features, processed by a unidirectional LSTM to model temporal connections. After RMS normalization, self-attention, and dropout, a sigmoid-activated linear layer predicts depressive or non-depressive text, enhancing performance in noisy contexts 1.

CNN: Captures local semantic features in code-mixed social media texts using Word2Vec embeddings. A 1D convolutional layer, max-pooling, and sigmoid output enable binary depressive text classification, performing well in short, noisy texts.

LSTM: Learns long-term dependencies in code-mixed text using GloVe embeddings. With memory gates and a sigmoid output, it classifies depressive content, suitable for sequential analysis in social media data.

BiLSTM: Processes code-mixed text bidirectionally with FastText embeddings, enhancing context. A BiLSTM layer and sigmoid output enable effective binary depressive text classification, excelling in emotion-rich content.

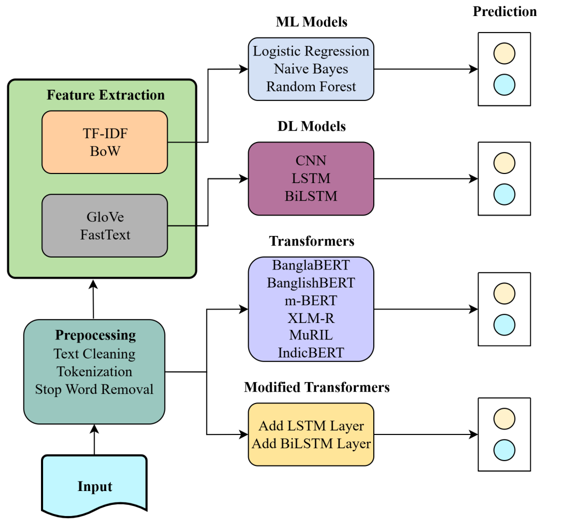

Figure 1: Proposed Methodology for Depression Detection.

BanglishBERT: Pre-trained on Bangla-English data, BanglishBERT captures informal text nuances. Fine-tuned with dropout 0.3, it achieves high accuracy in binary depressive text classification in noisy settings.

BanglaBERT: Monolingual BERT for Bangla, fine-tuned with dropout 0.3, provides robust depression detection in code-mixed texts, leveraging its 12-layer architecture for strong binary classification performance.

XLM-RoBERTa: Multilingual model pre-trained on 100+ languages, fine-tuned with dropout 0.3, excels in cross-lingual code-mixed data, enabling accurate binary classification of depressive sentiments.

MuRIL: Optimized for Indian languages, MuRIL supports mixed-script text. Fine-tuned with dropout 0.3, it effectively detects depressive phrases in code-mixed social media for binary classification.

mBERT: Pre-trained on 104 languages, mBERT generalizes across Bangla-English texts. Fine-tuned with dropout 0.3, it reliably classifies depressive content in code-mixed social media data.

IndicBERT: Lightweight ALBERT-based model for Indian languages, fine-tuned with dropout 0.3, efficiently classifies depressive text in Bangla-English data, ideal for resource-constrained environments.

BanglishBERT+LSTM: Combines frozen BanglishBERT embeddings with LSTM for temporal modeling. Fine-tuned with dropout 0.3, it excels in binary depressive text classification in noisy data.

Table 3: Hyperparameter Tuning.

	cc
Hyperparameter	**Value**
Data Split	
{60,70,80}	80-10-10
Dropout Rate	
{0.1, 0.2}	0.2
Batch Size	
{4–32}	8
Weight Decay	
{0.01, 0.001}	0.01
Epochs	
{5–20}	15
Patience	
{4, 8}	8
Learning Rate	
{(1–5)e-5}	3e-5

5 Results and Analysis

Table 4 presents a detailed comparison of several transformer-based backbones: BanglaBERT, BanglishBERT, and MuRIL. These are combined with three types of neural sequence models: LSTM, BiLSTM, and CNN. The focus is on detecting depressive text in Bangla-English code-mixed data.

Among all combinations, the BanglishBERT + LSTM model achieved the highest overall accuracy (0.8889) and the highest macro-averaged F1 score (0.8876). This indicates strong performance in identifying both depressive and non-depressive text. Its high precision (0.8915) and recall (0.8859) further show that it is a reliable and well-generalized model. It effectively handles the nuances of informal, bilingual expressions.

The BanglaBERT + BiLSTM and BanglaBERT + CNN models also showed competitive results. They had accuracy scores of 0.8718 and 0.8675, respectively. However, their performance is slightly behind the BanglishBERT-based configurations, particularly regarding precision and F1 score. Notably, MuRIL + CNN also achieved 0.8718 accuracy, which shows that CNN-based decoding layers can deliver strong outcomes when used with multilingual models.

Table 4: Performance of Models.

ML Models				
Classifier	**Pr(%)**	**Re(%)**	**F1(%)**	**Ac(%)**
Logistic Regression	0.82	0.81	0.82	**0.81**
Naive Bayes	0.81	0.81	0.81	0.80
SVM	0.79	0.78	0.79	0.77
Random Forest	0.69	0.65	0.74	0.67
DL Models				
Classifier	**Pr(%)**	**Re(%)**	**F1(%)**	**Ac(%)**
BiLSTM	82.00	82.00	82.00	**82.00**
CNN	79.00	80.00	79.00	80.00
LSTM	75.00	72.00	73.00	76.00
Transformers				
Classifier	**Pr(%)**	**Re(%)**	**F1(%)**	**Ac(%)**
BanglishBERT	84.86	83.94	84.21	**84.52**
BanglaBERT	83.49	82.30	82.60	83.00
XLM-RoBERTa	82.17	82.32	82.23	82.34
MuRIL	84.42	84.70	84.46	**84.52**
Hybrid Models				
Classifier	**Pr(%)**	**Re(%)**	**F1(%)**	**Ac(%)**
BanglishBERT+LSTM	89.15	88.59	88.76	**88.89**
BanglishBERT+BiLSTM	86.72	86.36	86.48	86.61
BanglishBERT+CNN	87.64	86.77	86.99	87.18
BanglaBERT+LSTM	86.38	86.49	86.42	86.47
BanglaBERT+BiLSTM	87.27	86.95	87.07	87.18
BanglaBERT+CNN	86.88	87.04	86.75	86.75
MuRIL+LSTM	83.38	83.49	83.42	83.48
MuRIL+BiLSTM	85.38	85.50	85.42	85.47
MuRIL+CNN	87.10	87.15	87.12	87.18

5.1 Error Analysis

A comprehensive quantitative and qualitative error analysis is conducted to provide detailed insights into the proposed model's performance.

5.1.1 Quantitative Analysis

The last row of Table 5 shows a misclassification example. Here, the model mistakenly labels a depressive text as non_depressive. The sentence expresses strong criticism and frustration toward systemic issues, such as business syndicates and government accountability.

5.1.2 Qualitative Analysis

Table 5 presents some predicted outputs of the proposed model. In the first and second texts, the model successfully predicted the depressive intent. However, in the third text, it misclassified a depressive sentence as non_depressive. This indicates that while the model captures most depressive cues,

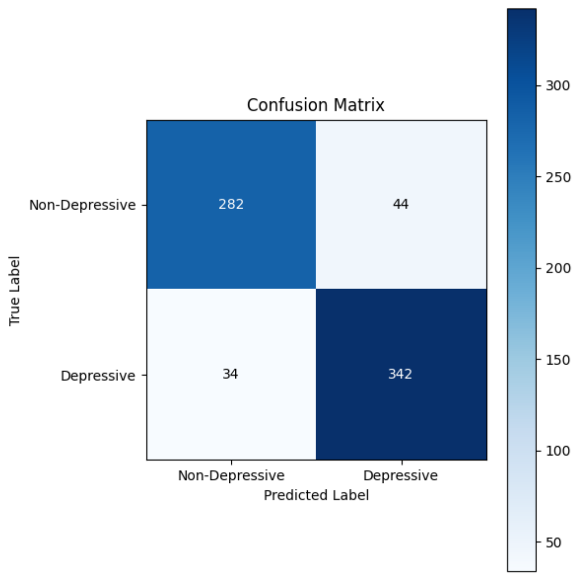

Figure 2: Confusion matrix of the proposed model.

it sometimes struggles with context-driven criticisms, particularly when expressions are subtle or indirect.

Table 5: Selected Examples for Qualitative Error Analysis.

Text	**Actual**	**Predicted**
Grame snake bite korle akhono manush ojha ar kache jay poison remove korte!	depressive	depressive
Eta sotti, jodi amra khabar kheye bachte na pari, tahole development diye ki hobe?	depressive	depressive
Bebsayi der syndicate er jonno amader government dayi.	depressive	non depressive

5.2 Model Interpretability Using LIME

To understand and explain the model's predictions, we used LIME (Local Interpretable Model-agnostic Explanations). LIME helps identify which specific words in a text most influenced the model's decision. This improves transparency and offers clarity, which is especially important in sensitive tasks like depression detection. It generates easy-to-understand explanations for individual predictions by locally approximating the model's behavior.

Example 1:
Input: "Bangladesher private company te job er nirdisto kono work hour nai, shuru ache kintu shesh nai somoyer."
Prediction: Depressive

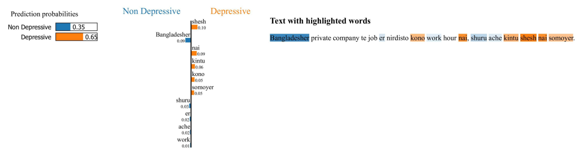

Figure 3: LIME Explanation for Example 1: Depressive Prediction.

Example 2:
Input: "Ami khubi depressed amader department ar team niye. Ato baje khele kivabe ora!"
Prediction: Depressive

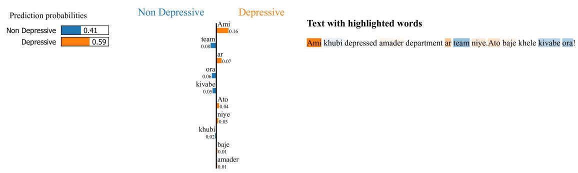

Figure 4: LIME Explanation for Example 2: Depressive Prediction.

6 Conclusion

In this study, we addressed depressive text detection in Bangla-English code-mixed data using machine learning, deep learning, and transformer models. While traditional methods achieved 81–82% accuracy, transformers improved results to 82–85%. Our hybrid approach, notably BanglaBERT+LSTM, achieved the best performance with 88.89% accuracy. These results highlight the effectiveness of combining contextual embeddings with recurrent layers, providing valuable benchmarks for future research in code-mixed sentiment analysis and culturally sensitive mental health monitoring.

6.1 Limitations

This study faces several limitations: (i) the dataset size (7,019 samples) is relatively small and may not capture the full diversity of Bangla-English code-mixed texts (ii) the limited data increases the risk of overfitting, reducing robustness on unseen data and (iii) subtle, context-dependent depressive expressions are often missed, limiting cultural and linguistic coverage.

6.2 Future Work

Future research can address these gaps by: (i) expanding the dataset with more diverse samples from multiple platforms (ii) extending from binary to multi-class classification for finer-grained depression severity detection and (iii) incorporating multimodal signals such as audio and video for richer, context-aware depression detection.

References

Kalika Bali, Jatin Sharma, Monojit Choudhury, and Yogarshi Vyas. 2014. Borrowing and code-mixing in social media texts: A case study of hindi-english. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 116–126.

Utsab Barman, Amitava Das, Joachim Wagner, and Jennifer Foster. 2014. Code mixing: A challenge for language identification in the language of social media. In *Proceedings of the First Workshop on Computational Approaches to Code Switching*, pages 13–23.

A Bhattacharjee et al. 2021. Banglabert: Language model pretraining and benchmarks for low-resource language understanding evaluation in bangla. *arXiv preprint arXiv:2101.00204*.

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 ACL*, pages 8440–8451.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186.

M Jahan, I Ahamed, MR Bishwas, and S Shatabda. 2019. Abusive comments detection in bangla-english code-mixed and transliterated text. In *Proceedings of the 2nd International Conference on Innovation in Engineering and Technology (ICIET)*. IEEE.

M Kerasiotis, L Ilias, and D Askounis. 2024. Depression detection in social media posts using transformer-based models and auxiliary features. *Social Network Analysis and Mining*, 14(1):196.

S Khanuja et al. 2021. Muril: Multilingual representations for indian languages. *arXiv preprint arXiv:2103.10730*.

Soumil Mandal, Amitava Das, and Björn Gambäck. 2020. Emotional analysis of bangla-english code-mixed data using deep learning. In *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT)*, pages 93–100.

TF Mumu, IJ Munni, and AK Das. 2021. Depressed people detection from bangla social media status using lstm and cnn approach. *Journal of Engineering Advancements*, 2(01):41–47.

MN Raihan, D Goswami, A Mahmud, A Anastasopoulos, and M Zampieri. 2023. Sentmix-3l: A bangla-english-hindi code-mixed dataset for sentiment analysis. *arXiv preprint arXiv:2310.18023*.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. “why should i trust you?” explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144.

B Sultana and KA Mamun. 2024. 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)*, pages 652–657. IEEE.

Sayema Sultana, Maisha Jahan, and Md Rezaul Karim. 2021. Sentiment analysis of bangla-english code-mixed data using traditional and deep learning methods. In *Proceedings of the International Conference on Bangla Speech and Language Processing (ICBSLP)*.

M Tareq, MF Islam, S Deb, S Rahman, and A Al Mahmud. 2023. Data-augmentation for bangla-english code-mixed sentiment analysis: Enhancing cross linguistic contextual understanding. *IEEE Access*, 11:51657–51671.

AH Uddin, D Bapery, and ASM Arif. 2019. Depression analysis of bangla social media data using gated recurrent neural network. In *2019 1st International conference on advances in science, engineering and robotics technology (ICASERT)*, pages 1–6. IEEE.