Bengali Emotion Classification Using Hybrid Deep Neural Network

Conference Paper · November 2023

DOI: 10.1109/AIKIIE60097.2023.10389834

2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE)

Bengali Emotion Classification Using Hybrid Deep Neural Network

1st Rezaul Haque
Dept. of CSE
East West University
Dhaka 1212, Bangladesh
rezaulh603@gmail.com

2nd Md Babul Islam
Dept. of DIMES
University of Calabria
Calabria, Italy
babulcseian@gmail.com

3rd Parameshachari B.D
Dept. of ECE
Nitte Meenakshi Institute of Technology
Bengaluru, India

4th Katura Gania Khushbu
Dept. of CSE
East West University
Dhaka 1212, Bangladesh
katurakhushbu@gmail.com

5th Shafiur Rahman
Dept. of Software
Engineering Daffodil International
University
Dhaka 1216, Bangladesh
shafiur016@gmail.com

6th Awan Ur Rahman
ABDRH
KUET
Khulna Bangladesh
awanrahman55@gmail.com

7th Md Helal Hossen
Dept. of CSE
University of Texas at Dallas
USA
helal.h017@gmail.com

8th Tanbin Islam Rohan
ABDRH
KUET
Khulna Bangladesh
tanbinirohan@gmail.com

***Abstract*—Emotion classification holds significant importance in various domains. However, the development of accurate emotion classification models for the Bengali language has been relatively limited, despite its vast speaker base. The unique characteristics of Bengali present several challenges for emotion classification. Consequently, there is an urgent demand for robust and contextually-aware emotion classification models tailored to the linguistic nuances of Bengali. This paper presents a comprehensive study on emotion classification in Bengali text, aiming to develop robust and effective models specific to the language. We explored a range of ML and DL models, including LR, SVC, CNN, LSTM, and BiLSTM. Additionally, we proposed novel hybrid architectures, combining CNN with LSTM and CNN with BiLSTM, to leverage both local and contextual information from Bengali text. However, the lack of comprehensive Bengali emotion datasets further hinders the development of dedicated emotion classification models for the language. To facilitate research, we created the 'Bengali Emotion Dataset' consisting of 14,334 social media comments, accurately labeled into seven emotion classes. The results demonstrate that the hybrid models, particularly CNN+BiLSTM, outperform individual ML and DL models, achieving the highest accuracy and F1 score of 88.45% and 88.42% respectively. The benchmark dataset and the success of the hybrid model pave the way for more empathetic and contextually-aware natural language processing applications for Bengali speakers.**

Keywords—emotion classification, hybrid deep learning, text classification

I. INTRODUCTION

Emotions are important to human interaction, influencing our perceptions, judgments, and behaviors. Emotion classification involves categorizing emotions expressed in text, has piqued the interest of researchers in recent years due to its im- portance in a variety of applications such as customer feedback analysis [1], human-computer interaction, and psychological support systems [2]. Effective emotion categorization models can offer insightful information about the emotional states of people and groups, providing a better understanding of their needs. Significant progress has been made in categorizing emotions for major languages like English. However, limited research resources have hindered the advancement of emotion categorization in smaller languages such as Bengali [3]. It is important to note that Bengali is one of the most widely spoken languages worldwide, with approximately 230 million native speakers residing primarily in Bangladesh and the Indian states of West Bengal, Tripura, and Assam [4], [5]. Despite its vast speaker base, the development of accurate emotion classification models for Bengali text have been relatively limited. Existing emotion classifiers for Bengali often suffer from limited accuracy and robustness, primarily due to challenges such as morphological complexity, contextual dependencies, limited linguistic resources, code-mixed language usage, out- of-vocabulary words, and linguistic diversity [6]. Therefore, there is an urgent demand for dedicated accurate emotion classification models that are tailored to the unique characteristics of the Bengali language.

One of the main objectives of this work is to offer a trustworthy and useful model for Bengali language emotion classification. Our comprehensive research methodology is illustrated in Figure 1. We used a variety of Machine Learning (ML) classifiers, including Stochastic Gradient Descent (SGD), Support Vector Classifier (SVC), and Logistic Regression (LR). Furthermore, Long Short-Term Memory networks (LSTM), Bidirectional LSTM (BiLSTM), and Convolutional Neural Networks (CNN) were among the Deep Learning (DL) approaches that we examined for their capabilities. We also considered the idea of hybrid DL algorithms that combine CNN with LSTM (CNN+LSTM) and CNN with BiLSTM (CNN+BiLSTM) in order to make use of the benefits of both architectures in extracting local and contextual information from Bengali text data. The research also intends to create the 'Bengali Emotion Dataset,' which includes 14,334 social media comments and is a complete dataset of Bengali emotions. These comments are labeled into one of the 7 emotion classes, namely joyful, sad, angry, disgusted, surprised, frightened, or none. Moreover, extensive evaluations were conducted to assess the proposed model's

2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE) | 979-8-3503-1646-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/AIKIIE60097.2023.10389834

performance and compare it with existing state-of-the-art emotion classification approaches.

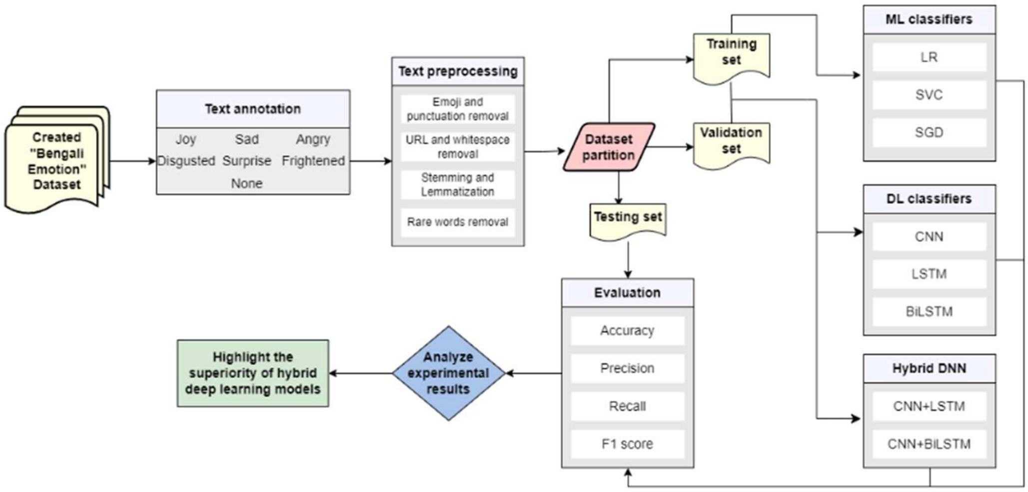

Fig. 1. Overall research methodology.

Through the proposed hybrid deep neural network model and the benchmark dataset, we aim to facilitate research in emotion analysis for Bengali text and foster the development of more empathetic and contextually-aware NLP applications for Bengali speakers. The key contributions of this research can be summarized as follows:

- Created a dataset of 14,334 comments with precise labeling into 7 emotion classes, enabling future Bengali emotion analysis.
- Introduced hybrid deep neural network models, enhancing accuracy by capturing spatial and temporal dependencies.
- Performed comparative analysis, assessing ML, DL, and hybrid DL models on the Bengali Emotion Dataset, offering insights into Bengali emotion classification challenges.

The remainder of this paper is structured as follows: Section 2 reviews related studies on Bengali emotion classification. In Section 3, we detail our proposed methodology, covering data collection, pre-processing, and hybrid model construction. Section 4 presents our experimental findings and offers a comprehensive discussion. Finally, in Section 5, we conclude the paper and suggest potential avenues for future research.

II. Related Works

The majority of emotion classification research has primarily focused on English and a few other widely spoken languages [7]. In [8] authors detect multi-class emotions from Bangla text using multinomial naïve bayes classifier along with various features such as stemmer, POS tagger, n-grams, Term Frequency Inverse Document Frequency (TF-IDF). The study showcased promising result of 78.6% accuracy in detection of emotions of texts written in three emotion classes (happy, sad and angry), thus underlining the viability of employing ML techniques within this context. A notable study conducted by Khan et al. [9] delved into emotion detection in Bangla text through the utilization of ML techniques. They employed SVC to detect five different emotions in Bengali text. Their approach achieved an accuracy of 62% on a dataset of 63,000 comments of fans towards a celebrity. Authors in [10] gathered and annotated a text corpus comprising user comments sourced from various Facebook groups, focusing on socio-economic and political issues. Their objective was to extract the fundamental emotions, including sadness, happiness, disgust, surprise, fear, and anger, conveyed through these comments. Notably, their most successful model was SVM, utilizing a non-linear RBF kernel. This model achieved an overall average accuracy score of 52.98% and an F1 score (macro) of 0.3324.

DL approaches have also been investigated for emotion classification in Bengali text. Emon et at. [11] addresses the prevalent issue of abusive content in online platforms in Bangladesh and aims to detect various types of abusive Bengali text using ML and DL algorithms. Their RNN-based algorithm emerges as the most effective, achieving the highest accuracy of 82.20% in detecting abusive text. This finding indicates the superiority of DL methods over traditional ML techniques for this specific task. Additionally, the exploration of advanced DL architectures and techniques, such as LSTM, and Transformers, holds the potential to enhance the performance of Bangla emotion detection models. While previous studies provide valuable insights, it did not delve into the specifics of emotion classification or explore hybrid architectures. As a result, authors [12] proposed a hybrid approach that integrated lexical features, syntactic features, and a hybrid deep neural network model combination of CNN and LSTM, called CLSTM. Their model exhibited an impressive accuracy score of 85.8% on a dataset of 42,036 Bengali comments divided into 4 classes. However, they have limited emotion categories but emotions are intricate and multifaceted phenomena that encompass a wide range of subtle nuances and variations.

Expanding the number of emotion classes in Bangla text data is crucial for enhancing the accuracy, cultural relevance, and overall effectiveness of emotion detection in the Bangla language. While some prior research has contributed to emotion analysis in Bengali and other languages, the focus on comprehensive emotion classification models specific to Bengali has been limited. Our work addresses this gap by proposing and evaluating a range of ML, DL, and hybrid models on a curated Bengali emotion dataset.

III. Methodology

A. Dataset Creation

The 'Bengali Emotion Dataset' is a comprehensive collection of 14,334 social media comments in the Bengali language, each labeled with emotions like happy, sad, anger, disgust, surprise, fear, or none. It was created by combining three publicly available datasets related to emotion detection in Bengali texts. Number of comments in each category of emotion is depicted in figure 2. The first dataset, 'Bangla YouTube Sentiment and Emotion [11],' contributed 2,900 comments from Bangla YouTube videos, providing diverse emotions expressed in YouTube comments. The second dataset, 'BanglaEmotion [13],' added 6,328 comments from various social media platforms, encompassing a broader range of emotions. The third dataset, 'BEmoC [14],' further enriched the dataset with 7,000 comments having various emotional labels. By amalgamating these datasets, the resulting 'Bengali Emotion Dataset' became a valuable resource for emotion detection models, offering a diverse collection of emotions prevalent in Bengali language texts.

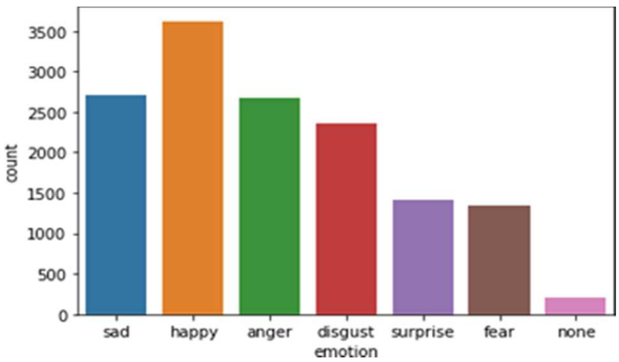

Fig. 2. Number of data per category.

B. Text Pre-processing

In the text pre-processing phase, we applied a series of techniques to clean and transform the raw text data from the 'Bengali Emotion Dataset' into a format suitable for emotion classification. These techniques included removing emojis, eliminating punctuation marks, getting rid of URLs, removing extra spaces and whitespaces, eliminating common stop words, and performing stemming and lemmatization to reduce words to their root forms. Additionally, words with low occurrence frequencies were removed to avoid overfitting. By employing these techniques, we aimed to enhance the overall performance and effectiveness of the emotion classification models on the Bengali text data. Figure 3 shows the most frequently occurring words in each category after text pre-processing.

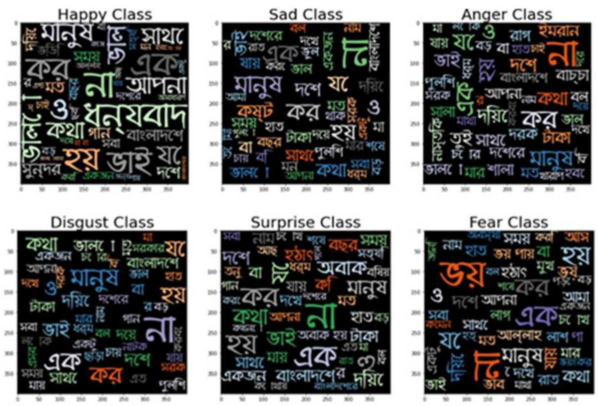

Fig. 3. Wordcloud representation of each category of emotions.

C. Feature Extraction

1) *TF-IDF:* TF-IDF is a widely used method in NLP to convert text documents into numerical feature vectors, enabling the training of ML classifiers [15], [16]. By using TF-IDF, we aim to capture the distinguishing characteristics of each document and create a feature matrix suitable for training ML classifiers [17]. The TF-IDF algorithm considers two essential components: Term Frequency (TF) and inverse document frequency (IDF). In the analysis of a specific document, the term frequency concept comes into play. It entails determining how often each word (term) appears within that document. By calculating the ratio of its occurrences to the total number of words in said document, this measure precisely assesses the significance of a word, shown in equation 1. Next, the inverse document frequency takes into account the global significance of each term across the entire corpus. Using equation 2, it penalizes common terms that appear in many documents and assigns higher weights to rare terms that carry more discriminatory power. Finally, the TF-IDF score for each term in a document is obtained by multiplying its TF value with its IDF value, shown in equation 3.

$$TF(a,b) = \frac{Term\ a\ frequency\ in\ document\ b}{Total\ words\ in\ document} \quad (1)$$

$$IDF(a) = log_2\left(\frac{Total\ documents}{documents\ with\ term\ a}\right) \quad (2)$$

$$TFIDF(a) = TF(a,b) \times IDF(a) \quad (3)$$

2) *Glove:* DL classifiers use numerical vectors instead of text to process words effectively. Word embedding, like GloVe, converts words into feature vectors, capturing their semantic representations. GloVe's log-bilinear regression model, based on word co-occurrence matrices, captures both local and global co-occurrence aspects, creating high-dimensional vector spaces. Unlike word2vec, GloVe explicitly explores co-occurrence matrix features for semantic similarity. With word pair associations, GloVe enhances the neural network with additional meanings, improving vector representations and classifier accuracy. Pretrained word vectors of various dimensions are available in GloVe. In our study, a vocabulary of size 41,322 and word vectors of 300 dimensions were used to create a GloVe embedding matrix for training DL and hybrid classifiers.

D. Training Models

1) *ML Models:* LR, SVC, and SGD classifiers were employed to identify the highest performing baseline model for the classification task. These algorithms are widely used for multiclass text classification. The models were trained with an 80% training data split, and the remaining 20% was used for testing. LR estimates class probabilities and assigns the most probable class as the predicted emotion label. It utilized L2 (Ridge) regularization with a regularization strength of 1.0 to prevent overfitting. The optimization algorithm used for training was Limited-memory Broyden–

Fletcher–Goldfarb–Shanno (LBFGS), with a maximum of 100 iterations allowed for convergence. SGD is an iterative optimization algorithm efficient for training linear classifiers on large datasets. We used multiclass log loss for multiclass classification. The learning rate was set to 'optimal' to adapt based on data characteristics and iterations, with a maximum of 1000 iterations allowed for convergence. SVC-rbf is a non-linear classifier using the radial basis function kernel, suitable for complex data distributions. The regularization parameter was set to 1.0, and the gamma parameter was determined automatically based on feature count. SVC-linear is a linear classifier appropriate for linearly separable data, using a linear hyperplane. For training, the regularization parameter was set to 1.0, balancing margin maximization and error minimization.

2) DL Models: To outperform the performance of baseline model we explored DL algorithms such as CNN, LSTM, and BiLSTM. All the DL algorithms were trained on a batch size of 128, a learning rate of 0.001, and 100 epochs. In the context of text classification, CNN utilizes convolutional layers to capture local patterns and features within the input text. In our study, we employed a 1D CNN architecture to process the one-dimensional text data. The model used convolutional layers with filters to extract local features and patterns, followed by max-pooling layers to down-sample features. The final output passed through fully connected layers for classification. For training, we used an embedding dimension of 100 and filter sizes of [3, 4, 5] to capture various n-grams. Each filter size had 128 filters, and we applied a 0.5 dropout rate to mitigate overfitting.

LSTM is a type of RNN designed to overcome the vanishing gradient problem in traditional RNNs [18]. It can capture long-range dependencies in sequential data, making them effective for understanding the emotional context and temporal dynamics of Bengali text. We used LSTM layers to process the sequential text data and extract contextual information. We set the embedding dimension to 100 and employed 128 LSTM units with a 0.5 dropout rate to prevent overfitting. In addition, we utilized BiLSTM for bidirectional processing, allowing the model to capture context from both directions simultaneously. This approach enhances the model's ability to understand the nuances of emotions expressed in Bengali text. For BiLSTM, we maintained the same embedding dimension and 128 LSTM units, along with a 0.5 dropout rate to avoid overfitting.

3) Hybrid Models: Hybrid models were trained on a 70% training set and evaluated on a 10% validation set for performance monitoring and overfitting prevention. The remaining 20% of the dataset served as a separate testing set to evaluate generalization on unseen data. A batch size of 256 was used, and a learning rate of 0.001 determined the step size for parameter adjustments during 100 training epochs. The CNN+LSTM model is a hybrid architecture that incorporates both CNN and LSTM components. The combination of CNN and LSTM allows the model to extract relevant features from individual words while also understanding the overall context and long-range dependencies present in the text [12].

The CNN layer is responsible for extracting relevant features from the input sequences with a shape of $(batch\ size, time\ steps, input\ features)$. The output of the CNN layer is represented as CNN_output and the CNN layer as CNN in equation 4.

$$CNN_{output} = CNN(input\_sequences) \tag{4}$$

These features are passed into LSTM layers to capture sequential dependencies and context in the text. The LSTM equations (5-9) involve (X_t) as the input at time step (t), (H_{t-1}) as the previous hidden state, (W) and (b) for weight and bias parameters, (σ) representing the sigmoid function, and $(\tanh)$ as the hyperbolic tangent function. With 7 output categories, the Dense layer (C) was set to 7. The output layer parameters are$(W_{output})$for weights, (b_{output}) for biases, and (P) for class probabilities generated by the softmax function. For training, the model had an embedding dimension of 100 for word embeddings. The CNN component used filter sizes [3, 4, 5] to capture various n-grams with 128 filters per size. The LSTM component had 128 LSTM units, and a 0.5 dropout rate was applied to prevent overfitting during training.

$$i_t = \sigma(W_{xi} \cdot X_t + W_{hi} \cdot H_{t-1} + b_i) \tag{5}$$

$$f_t = \sigma\left(W_{xf} \cdot X_t + W_{hf} \cdot H_{t-1} + b_f\right) \tag{6}$$

$$o_t = \sigma(W_{xo} \cdot X_t + W_{ho} \cdot H_{t-1} + b_o) \tag{7}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh(W_{xc} \cdot X_t + W_{hc} \cdot H_{t-1} + b_c) \tag{8}$$

$$H_t = o_t \cdot \tanh(c_t) \tag{9}$$

$$P_{\text{CNN+LSTM}} = \text{softmax}\left(W_{\text{output}} \cdot H_t + b_{\text{output}}\right) \tag{10}$$

The CNN+BiLSTM model is another hybrid architecture that combines CNN and BiLSTM components. Like the CNN+LSTM model, it begins by using CNN layers to extract local features and patterns from the input text. The equations are similar to LSTM but with two sets of weights and states for the forward and backward passes. The 'Output Layer' shown in equation 14, takes the information captured by the CNN and BiLSTM layers and maps it to a probability distribution over the different classes using the softmax function. For training, the model is configured with an embedding dimension of 100, representing the size of the word embeddings used for word representation. The CNN component uses filter sizes of [3, 4, 5], and 128 filters are applied to each filter size. The BiLSTM component consists of 128 BiLSTM units, and a dropout rate of 0.5 is applied to mitigate overfitting during training.

$$H_t^{fwd} = LSTM^{fwd}\left(X_t, H_{t-1}^{fwd}\right) \tag{11}$$

$$H_t^{bwd} = LSTM^{bwd}(X_t, H_{t+1}^{bwd}) \tag{12}$$

$$H_t = \left[H_t^{fwd}, H_t^{bwd}\right] \tag{13}$$

$$P_{\text{CNN+BiLSTM}} = \text{softmax}\left(W_{\text{output}} \cdot H_t + b_{\text{output}}\right) \tag{14}$$

E. Evaluation

We assessed the performance of classifiers using traditional multi-class evaluation metrics. In the multi-class scenario, each unit receives a predicted probability value. We calculated macro-average precision and recall for the generic class (K), as the arithmetic mean of precision and recall metrics for individual classes (equations 17 and 18). The Macro F1-Score (equation 19) is the harmonic mean of Macro-Precision and Macro-Recall. Using macro-average methods, we calculated an overall mean of various measures. This approach gives equal weight to classes of different sizes, making the evaluation independent of class size. A high Macro-F1 value indicates strong overall performance, while low values suggest poorly predicted classes.

$$Precision_K = \frac{TP_K}{TP_K + FP_K} \quad (15)$$

$$Recall_K = \frac{TP_K}{TP_K + FN_K} \quad (16)$$

$$MacroPrecision = \frac{\sum_{K=1}^{K} Precision_K}{K} \quad (17)$$

$$MacroRecall = \frac{\sum_{K=1}^{K} Recall_K}{K} \quad (18)$$

$$MacroF1 = \frac{2 * (acroPrecision * MacroRecall)}{MacroPrecision^{-1} + MacroRecall^{-1}} \quad (19)$$

IV. Result Analysis and Discusssion

A. *Performance of the Classifiers*

The TABLE I presents the performance metrics of various models used for emotion classification. Among the ML models, the SVC-rbf model achieved the highest accuracy of 85.86%. DL models, including CNN, LSTM, and BiLSTM, exhibited slightly better accuracy compared to the ML models for this emotion classification task. CNN, LSTM, and BiLSTM achieved accuracies of 86.15%, 86.98%, and 87.56%, respectively. The hybrid models, CNN+LSTM, and CNN+BiLSTM, obtained accuracies of 87.99% and 88.45%, respectively. Notably, the hybrid model CNN+BiLSTM demonstrated the highest accuracy, precision, recall, and F1 Score of 88.45%, 88.21%, 88.96%, and 88.42%, respectively among all the models, making it the best-performing model.

TABLE I. Results of Experimental Emotion Classifiers.

Type	Model	Accuracy	Precision	Recall	F1
ML	SVC-rbf	0.8586	0.8575	0.8678	0.8626
	SVC-linear	0.8513	0.8503	0.8605	0.8554
	SGD	0.8505	0.8494	0.8596	0.8545
DL	CNN	0.8615	0.8603	0.8745	0.8610
	LSTM	0.8698	0.8616	0.8798	0.8654
	BiLSTM	0.8756	0.8720	0.8799	0.8796
Hybrid	CNN+LSTM	0.8799	0.8778	0.8801	0.8801
	CNN+BiLSTM	0.8845	0.8821	0.8896	0.8842

The results show that combining CNN and BiLSTM architectures is highly effective for accurate emotion classification. Figure 4 confirms that hybrid models consistently outperformed individual DL models, underscoring the advantages of combining various DL techniques for superior emotion classification performance. The hybrid algorithms' success can be attributed to their ability to effectively capture local and contextual information, their contextual awareness, improved feature extraction, reduced overfitting, and the combination of complementary strengths from different architectures. These factors collectively contribute to the superior performance of hybrid models over individual ML or DL models in the context of the Bengali language.

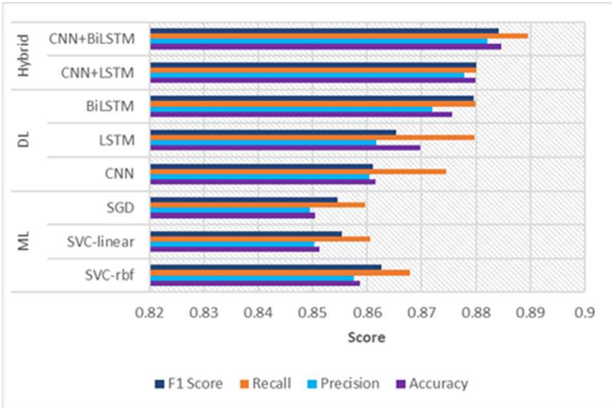

Fig. 4. Classification score of each classifier.

B. *Result Validation*

The confusion matrix shown in figure 5 represents the classification results of highest performing model CNN+BiLSTM for different emotions, including Happy, Sad, Anger, Disgust, Surprise, Fear, and None. Each row in the matrix represents the predicted emotion, while each column represents the true emotion. 'Happy' class has the lowest misclassification percentage of approximately 12.2%, indicating that the model performs relatively well in accurately classifying instances labeled as 'Happy.' On the other hand, 'Surprise,' 'Fear,' and 'None' have the highest misclassification percentages of around 25.1%, 23.4%, and 23.1%, respectively. These classes seem to be more challenging for the model to classify correctly, as they have the highest rates of false positives among all the classes.

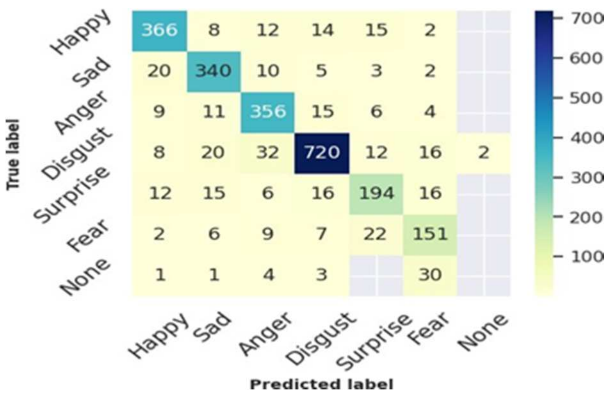

Fig. 5. Confusion matrix of CNN+BiLSTM.

The learning curve shown in figure 6 illustrates the performance of the CNN+BiLSTM over 100 epochs. Notably, both training and validation accuracy consistently increased from epoch 0 to 100, indicating effective learning and generalization. Training and validation loss steadily decreased, demonstrating the model's improved error minimization and emotion prediction. Overall, the learning curve analysis demonstrated the CNN+BiLSTM model's excellent learning capabilities and ability to avoid overfitting. The similar training and validation accuracies suggest that the model does not overfit to the training data, ensuring its reliability in real-world emotion classification tasks.

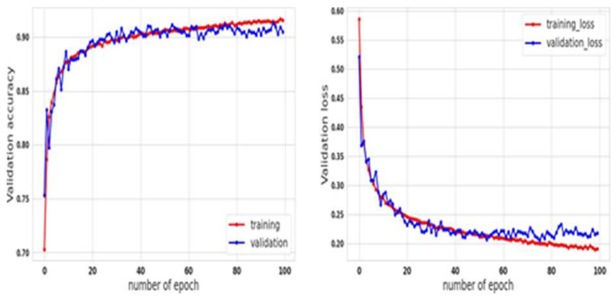

Fig. 6. Learning curve of CNN+BiLSTM.

C. State-of-the-art Comparison

TABLE II presents a comprehensive comparison of state-of-the-art Bengali emotion classification methods. Our approach stands out with an impressive 88.45\% accuracy, achieved with a smaller dataset of 14,334 comments, compared to [12]'s 42,036 and [9]'s 63,000 comments. Moreover, our CNN+BiLSTM model with seven classes excels in handling complex tasks, in contrast to [12]'s CNN+LSTM with four classes and [11]'s simpler RNN with seven classes. Notably, our approach outperforms SVM-based methods [9, 10].

In summary, our approach demonstrates remarkable accuracy, leveraging a smaller dataset and a complex task, highlighting the effectiveness of our chosen architecture and hybrid DL techniques.

TABLE II. COMPARISON WITH THE STATE-OF-THE-ART METHODS.

Reference	Number of data	Model	Classes	Accuracy
Our Approach	**14,334**	**CNN+BiLSTM**	**7**	**88.45%**
[12]	**42,036**	**CNN+LSTM**	**4**	**85.80%**
[11]	**4,700**	**RNN**	**7**	**82.20%**
[9]	**63,000**	**SVC**	**3**	**62%**
[10]	**6,314**	**SVC**	**6**	**52.98%**

V. CONCLUSION

Emotion classification plays a vital role in numerous applications, ranging from effective computing to mental health support systems, making accurate and contextually-aware models essential for understanding and responding to human emotions. However, the development of such models in Bengali has been relatively limited, despite its vast speaker base and significance. Our study focused on the development of robust emotion classification models tailored to the unique characteristics of the Bengali language. Our experimentation revealed that hybrid models, specifically CNN+LSTM and CNN+BiLSTM, outperformed individual ML and DL models. These hybrid architectures effectively captured both local and contextual information, enabling them to better understand the sequential nature of Bengali text and its emotional nuances. The combination of CNNs and LSTMs/BiLSTMs in hybrid models facilitated improved feature extraction, contextual awareness, and reduced overfitting, contributing to their superior performance in emotion classification. By combining hybrid models with the benchmark Bengali Emotion Dataset, researchers and practitioners can benefit from improved emotion classification performance in various NLP tasks for the Bengali language. Despite the promising results, there are still some challenges to address in emotion classification for Bengali text. While the model achieved remarkable accuracy overall, addressing the misclassification challenges for these specific emotions with the use of data augmentation and feature engineering could further improve its performance. Future research can expand the dataset, explore different hybrid architectures, and integrate multi-modal data for comprehensive emotion classification. Real-time emotion analysis and model adaptation to evolving linguistic patterns offer exciting avenues for investigation. Continued efforts in emotion analysis for Bengali can lead to more effective and contextually-sensitive models with broader practical applications.

REFERENCES

[1] R. Haque, P. K. Pareek, M. B. Islam, F. I. Aziz, S. Das Amarth, and K. G. Khushbu, "Improving Drug Review Categorization Using Sentiment Analysis and Machine Learning," 2023 International Conference on Data Science and Network Security (ICDSNS), pp. 1–6, Jul. 2023, doi: 10.1109/ICDSNS58469.2023.10245841.

[2] Jayaprakash JS, Balasubramanian K, Sulaiman R, Hasan MK, Parameshachari BD, Iwendi C. Cloud Data Encryption and Authentication Based on Enhanced Merkle Hash Tree Method. Computers, Materials & Continua. 2022 Jul 1;72(1).

[3] S. R. Titli and S. Paul, "Automated Bengali abusive text classification: Using Deep Learning Techniques," 2023 International Conference on Advances in Electronics, Communication, Computing and Intelligent In- formation Systems (ICAECIS), pp. 1–6, Apr. 2023, doi: 10.1109/ICAE- CIS58353.2023.10170294.

[4] Begum S, Banu R, Ahammed GA, Parameshachari BD. Performance degradation issues of PV solar power plant. In2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT) 2017 Dec 15 (pp. 311-313). IEEE.

[5] M. Hasan, L. Islam, I. Jahan, S. M. Meem, and R. M. Rahman, "Natural Language Processing and Sentiment Analysis on Bangla So- cial Media Comments on Russia–Ukraine War Using Transformers," Vietnam Journal of Computer Science, pp. 1–28, May 2023, doi: 10.1142/S2196888823500021.

[6] Sumathi AC, Akila M, Pérez de Prado R, Wozniak M, Divakarachari PB. Dynamic bargain game theory in the internet of things for data trustworthiness. Sensors. 2021 Nov 16;21(22):7611.

[7] A. Majeed, H. Mujtaba, and M. O. Beg, "Emotion detection in roman urdu text using machine learning," Proceedings - 2020 35th IEEE/ACM International Conference on Automated Software En- gineering Workshops, ASEW 2020, pp. 125–130, Sep. 2020, doi: 10.1145/3417113.3423375.

[8] Bhaskar K, Mohan CK, Lingam M, Reddy PV, Venkateswarlu V, Rao MY. & Ravichandran, V.(2009). Development of SLN and NLC enriched hydrogels for transdermal delivery of nitrendipine. vitro, ex-vivo.:98-113.

[9] Md. S. S. Khan, S. R. Rafa, A. E. H. Abir, and A. K. Das, "Sentiment Analysis on Bengali Facebook Comments To Predict Fan's Emotions Towards a Celebrity," Journal of Engineering Advancements, vol. 2, no. 03, pp. 118–124, Jul. 2021, doi: 10.38032/JEA.2021.03.001.

[10] Datla R, Chalavadi V, Mohan CK. A framework to derive geospatial attributes for aircraft type recognition in large-scale remote sensing images. InFourteenth International Conference on Machine Vision (ICMV 2021) 2022 Mar 4 (Vol. 12084, pp. 172-179). SPIE.

[11] E. A. Emon, S. Rahman, J. Banarjee, A. K. Das, and T. Mittra, "A Deep Learning Approach to Detect Abusive Bengali Text," 2019 7th International Conference on Smart Computing and Communications, ICSCC 2019, Jun. 2019, doi: 10.1109/ICSCC.2019.8843606.

[12] R. Haque, N. Islam, M. Tasneem, and A. K. Das, "Multi-class sentiment classification on Bengali social media comments using machine learn- ing," International Journal of Cognitive Computing in Engineering, vol. 4, pp. 21–35, Jun. 2023, doi: 10.1016/J.IJCCE.2023.01.001.

[13] M. A. Rahman and Md. H. Seddiqui, "BanglaEmotion: A Benchmark Dataset for Bangla Textual Emotion Analysis," vol. 1, 2020, doi: 10.17632/24XD7W7DHP.1.

[14] MD. A. Iqbal, A. Das, O. Sharif, M. M. Hoque, and I. H. Sarker, "BEmoC: A Corpus for Identifying Emotion in Bengali Texts," SN Computer Science 2022 3:2, vol. 3, no. 2, pp. 1–17, Jan. 2022, doi: 10.1007/S42979-022-01028-W.

[15] M. B. Islam, S. Hasibunnahar, P. K. Shukla, P. K. Shukla, P. Rawat, and J. Dange, "Twitter Opinion Mining on COVID-19 Vaccinations by

Machine Learning Presence," Lecture Notes in Networks and Systems, vol. 479, pp. 37–55, 2023, doi: 10.1007/978-981-19-3148-2 4/COVER.

[16] M. B. Islam, S. Hasibunnahar, P. K. Shukla, P. K. Shukla, and V. Jain, "Pandemic Outbreak Time: Evaluation of Public Tweet Opinion by Machine Learning," Proceedings of 2022 IEEE International Conference on Current Development in Engineering and Technology, CCET 2022, 2022, doi: 10.1109/CCET56606.2022.10080351.

[17] R. Haque, S. H. Laskar, K. G. Khushbu, M. J. Hasan, and J. Uddin, "Data-Driven Solution to Identify Sentiments from Online Drug Reviews," Computers 2023, Vol. 12, Page 87, vol. 12, no. 4, p. 87, Apr. 2023, doi: 10.3390/COMPUTERS12040087.

[18] R. Haque, N. Islam, M. Islam, and M. M. Ahsan, "A Comparative Analysis on Suicidal Ideation Detection Using NLP, Machine, and Deep Learning," Technologies 2022, Vol. 10, Page 57, vol. 10, no. 3, p. 57, Apr. 2022, doi: 10.3390/TECHNOLOGIES10030057.