A Comparative Study of LSTM and Bi-LSTM Architectures with Attention for Bangla News Classification

Md. Shazedur Rahman[1], Parvaj Kazi[1], Mir Mynul Ahasan Mim[1], Sadman Sadik Khan[1*], Nitta Nando Roy[1], and Md Anisur Rahman[2]

[1] Daffodil International University, Dhaka, Bangladesh
[2] Jahangirnagar University, Dhaka, Bangladesh

{shazedur15-14056, mir15-13953} diu.edu.bd, sadman15-13696@diu.edu.bd*, nitta15-6336@s.diu.edu.bd
{parvajkazi01, career.mdanisurrahman} @gmail.com

Abstract. News classification is an important task to perform in NLP, more so when dealing with low-resource languages such as Bangla. However, Bangla comes with its own set of challenges like different morphology, complex syntax, and a very acute shortage of large annotated corpora. In this work, Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), LSTM with Attention, and Bi-LSTM with Attention were tested to classify Bangla news articles into four broad categories: sports, national, international, and entertainment. The balanced Kaggle dataset has 11,904 labeled samples. The dataset underwent preprocessing that included normalization, tokenization, padding, and removal of stopwords. Models were trained using the Adam optimizer for twenty epochs on categorical crossentropy loss. Results showed that Bi-LSTM+Attention set the highest validation accuracy of 96%, outperforming all other models. These results have paved the way that attention-based deep learning models can effectively classify Bangla news domains.

Keywords: LSTM, Bi-LSTM, Bangla News Classification, Natural Language Processing, Attention Mechanism, Computational Linguistics

1 Introduction

The last decade has witnessed rapid developments in NLP, with text classification being a primary task, including spam filtering, sentiment analysis, topic label assignments, and news categorization. Automatic text classification is necessary to curb the problem of tremendous information overload caused by an exponential increase in digital content. In the real world, news classification has relevance for media archiving, recommendation systems for personalized news delivery, and monitoring of emerging topics and public opinion trends. Among these, automatic monitoring systems for news classification take the lead for

M. S. Arefin et al. (eds.), *Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025)*, Advances in Intelligent Systems Research 206,
https://doi.org/10.2991/978-94-6239-664-7_31

media archiving and personalized news delivery. With the sudden upsurge for the digitalization of news in regional languages, the requirement for intelligent systems capable of processing content and accurate classification has become crucial.

Bangla is spoken by more than 250 million speakers worldwide and ranks among the top five native languages. Despite such prominence, Bangla lacks adequate representation in computational linguistics mainly due to a paucity of annotated resources, standard tools, and benchmarking datasets. Such a linguistic gap stands as a thick barrier in developing particularly strong NLP applications, especially for tasks requiring fine-grained linguistic treatment such as news classification.

The morphological structure of the language is much richer, the word formation rules much more complicated, and innumerable dialects exist while English has comparatively few such features and rules. This greatly increases the difficulties in attribute extraction and semantic interpretation in machine-learning methods. Models also tend to overfit and generalize poorly since a large-scale labeled Bangla corpus has yet to be compiled. Despite relatively limited success obtained with conventional machine models such as Naive Bayes and Support Vector Machines (SVMs), deep learning starts to show more prominent results in recent developments. For sequential data processing, RNNs, typically LSTMs and Bi-LSTMs, have grabbed the highest priority in NLP. They are model-intensive when it handles long-range dependencies; hence these are quite good to model complicated sentence structure properties intrinsic to Bangla. The problem arises when some parts of the input should be attended to more than others. This is where attention enters.

An Attention Mechanism allows the model to attend to some relevant parts of the text to increase interpretability and performance in sequence models.In this work, four models of LSTM, Bi-LSTM, LSTM with Attention, and Bi-LSTM with Attention are proposed and contrasted for Bangla news classification.The bulk of our work speaks for a deep comparative study of these deep learning models on the balanced and manually preprocessed Bangla news data set. We stress the gain in performance by implementing attention mechanisms and also show the way bidirectionality helps process contextual information that is peculiar to Bangla sentence constructs. Additionally, we provide a reproducible big clean pipeline and training setup that could be used as a baseline for further Bangla NLP works churned in the news categorization domain. This will provide a more sound comparative analysis using the balanced dataset together with standardized evaluation metrics and will therefore become a convincing baseline for further research.

2 Literature Review

Since recently, more focus is put by researchers on Bangla news classification into deep learning models for semantic and syntactical features of texts. Earlier methods used conventional machine learning models such as Naïve Bayes and Support Vector Machines relying on TF-IDF and Bag-of-Words representations [1][2]. The ease of implementation of machine learning representations has been overweighed by their inability to follow the Bangla language in contextual flows, hence their average level of performance.

With the rise of deep learning, researchers started leaning toward CNNs and RNNs. Siam et al. [3] introduced an LSTM-based classifier for Bangla news articles that surpassed SVM and Naïve Bayes on their own dataset. Roy et al. [4] showed that Bi-LSTM is more effective than LSTM, with accuracy above 91% on balanced categorizations. Hossain et al. [5] employed CNN and LSTM combined with word embeddings to classify Bangla news headlines.and LSTM in combination with word embeddings for Bangla news headline classification.

In their study, Hasan et al. [6] proposed a hybrid deep learning architecture by combining convolutional and Bi-LSTM layers and found it suitable for small-scale Bangla datasets. Potrika is a large Bangla news corpus released by Hossain et al. [7], facilitating numerous downstream tasks, including classification. As their study explains, the ability of such a model to generalize is greatly enhanced when pre-trained on massive corpora.

Islam et al. [8] initiated the discussion of different algorithms on Bangla news text, comparing LSTM, GRU, and Bi-GRU and concluding the best performance of Bi-GRU with attention mechanisms. Rahman et al. [9] and Nahar et al. [10] stressed the importance of custom embeddings and preprocessing of Bangla data. According to Sultana et al. [11] and Khan et al. [12], attention-based models are capable of improving F1 scores of the classification by 3-4% over the baselines.

Uddin et al. [13] have studied hierarchical attention networks and concluded they perform better with applications in news with long texts as context. Ahmed et al. [14] tested some Transformer models such as mBERT and XLM-R while achieving state-of-the-art results but observed the high computational cost involved. Talukder et al. [15] prepared a labeled dataset of Bangla headlines to test various deep models.Rashid et al. [16] had yet another perspective on the use of attention in summarization and classification; they stated that attention locals important keywords. Dey et al. [17], Rahman et al. [18] compared the word embeddings (the GloVe, FastText) with the LSTM models and concluded that GloVe gives some consistent improvements. Ahmed et al. [19] worked on Bangla text categorization with hybrid deep networks, while Rahman et al. [20] presented language-specific BiLSTM attention approaches to news classification.

One can argue that attention-based processes and transformers for their contextual understanding are prevailing in research-oriented Bangla news classification tasks. This argument is supported by our evaluation that compares four architectures-LSTM, Bi-LSTM, LSTM+Attention, and Bi-LSTM+Attention-on a balanced news dataset.

3 Methodology

The methodology employed in this study adheres to a systematic, multi-phase framework beginning with data acquisition and ending in comprehensive model evaluation. Each stage in the process is thoroughly depicted in the accompanying diagram. Below is an outline of each phase:

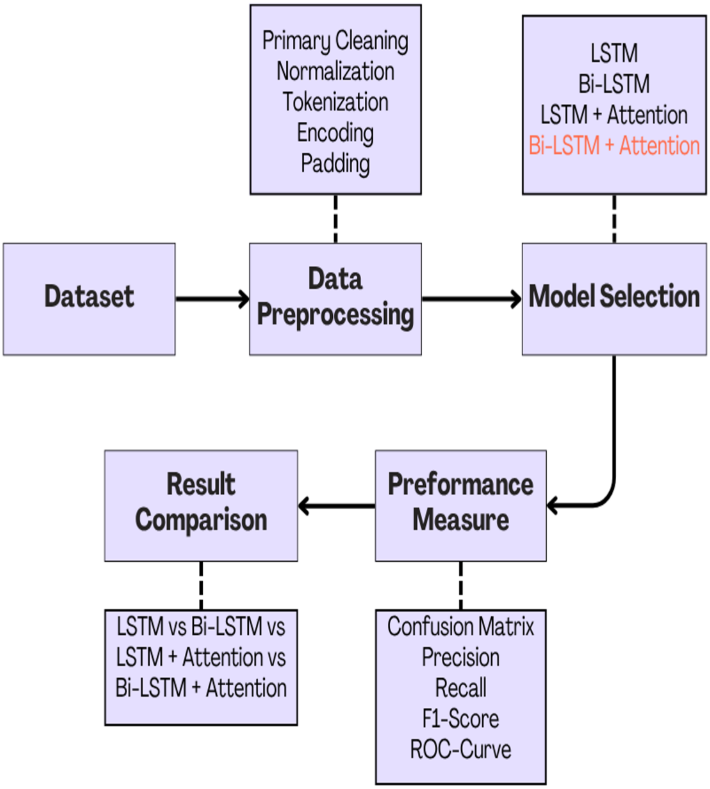

Fig. 1. Proposed Methodology

3.1 Data Collection

The dataset for this study was downloaded from Kaggle. This dataset is meant to classify Bangla news. It contained 11,904 labeled Bangla news items, sorted into four categories: National, International, Sports, and Entertainment. Each category has exactly 2,976 samples. Having such a distribution is very important to perform unbiased training of the model and to evaluate the performance metrics. A sample of the dataset was shown in Fig.2, whereas Fig.3 describes the data distribution among categories.

	title	published_date	reporter	category
0	সিন নদীতে ব্যাপক দূষণ, স্থগিত করা হলো ট্রায়াথ...	30th July, 2024 5:59 pm	আরআইএম	sports
1	এসিসির নতুন সভাপতি হচ্ছেন মহসিন নাকভি!	30th July, 2024 5:25 pm	NaN	sports
2	২০৩০ ও ২০৩৪ বিশ্বকাপের হোস্ট জানা যাবে দুইদিন পর!	30th July, 2024 1:22 pm	আরআইএম	sports
3	প্যারিস অলিম্পিক: কোয়ার্টার ফাইনালে ওঠার লক্ষ্...	30th July, 2024 11:54 am	এনকে	sports
4	আজ টিভিতে যা দেখবেন (৩০ জুলাই ২০২৪)	30th July, 2024 7:01 am	এনকে	sports

Fig. 2. Sample of the Dataset

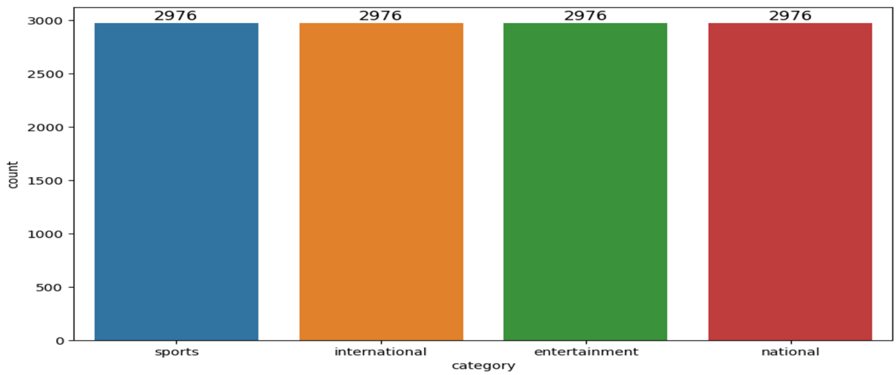

Fig. 3. Data Distribution

3.2 Data Preprocessing

During this phase, we applied the preprocessing steps to the Bangla news data to prepare it for deep learning classification models. The preprocessing operations included text normalization, tokenization, padding, and encoding to standardize the data and derive a suitable numerical representation. First, we checked the dataset for null values or missing data, but it was clean, with no null values present. Therefore, we applied normalization to texts to set all entries to lowercase, remove characters such as emojis or special symbols, eliminate several punctuations, and strip extra white spaces. For this task, we used a BNLP-based normalizer, which could work well with Bangla text. It also comes with stemmer and lemmatizer to bring the words down to their roots.

Initially, articles underwent tokenization through use of the Keras Tokenizer. Each article was divided into tokens during this phase of processing. There were 35,150 unique tokens identified in this step. After tokenization, a vocabulary was constructed using these tokens, and all sequences were padded to a fixed length of 121 tokens so they could be fed into the models. The classification labels were encoded using label encoding before any further manipulation was done. Since the dataset was balanced, there was no need for oversampling. Eventually, the dataset's original class distribution was visualized on a frequency graph for verification of equal representation of all four categories.

3.3 Model Selection

For a complete evaluation on deep learning approaches on the classification of Bangla news articles, four sequence models were implemented and compared, namely: LSTM, Bi-LSTM, LSTM with attention, and Bi-LSTM with attention. These models aimed to capture the temporal and contextual dependencies in Bangla news content. The first architecture that was designed is a normal Long Short-Term Memory (LSTM) network. In this model, there is one unidirectional LSTM layer that takes the input sequences from left to right. Long-Short Term memory can really pay homage to learning long-term dependencies present in text data and, therefore, is suitably fit for handling any sorts of sequences, be it a sentence or a paragraph.

By contrast, this paper incorporates the enhancement in temporal learning through using the Bidirectional LSTM (Bi-LSTM). The architecture constitutes two parallel LSTM layers that read the input sequence in a forward manner in one, and backward in the other, allowing the model to get a greater sense of context from both preceding words and following words. A bidirectional approach generally proves to be more effective for classification tasks in which the meaning of a word depends on the tokens near it.

The Attention mechanism was introduced to promote more relevant focus on the portions of each sequence by the models. In the LSTM + Attention model, the attention layer sits atop the LSTM layer, thereby allowing more weight to be assigned to relevant words in the input sequence in coming to a categorization decision. Lastly, the Bi-LSTM + Attention model took advantage of bidirectional temporal learning and attention. In other words, it put an attention mechanism upon a Bi-LSTM layer, whereby the model focused on key tokens and considered the context in both directions. This model yielded the highest performance measures among all the studied architectures.

Embedding was the first layer in all four models for word representation, followed by LSTM or Bi-LSTM layers, with dropout regularization at 0.3 to avoid overfitting, and a fully connected dense layer with softmax activation for final classification.

3.4 Model Training

All four models, i.e., LSTMs, BiLSTMs, LSTM with Attention, and Bi-LSTM with Attention, were trained using the Adam optimizer at a 0.001 learning rate. The dataset was split into training (80%) and testing (20%). Common to all models was the use of embedding layers followed by LSTM/Bi-LSTM and/or attention layers. Lastly, a dense output layer with softmax activation was used. Dropout was applied with a rate of 0.3 to avoid overfitting. Models were trained over 20 epochs with a batch size of 64. Categorical crossentropy was employed for the loss function, with accuracy as the key metric when training.

3.5 Model Evaluation

The different models were assessed by using standard classification metrics: Accuracy, Precision, Recall, and F1-Score. On the other hand, per-class performance was evaluated using a Confusion Matrix and an ROC-AUC Curve.

Precision: How many of the predicted positive instances are in fact true positive instances.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

Recall: How capable the model is of capturing all relevant (actual positive) instances.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

F1-Score: The harmonic mean of precision and recall; it balances false positives and false negatives.

$$F1 - score = 2 \times \frac{Precision * Recall}{Precision + Recall} \tag{3}$$

Accuracy: The overall proportion of instances that are correctly classified out of the given predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Where, TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives,

Confusion Matrix: This is a summary table showing correct and incorrect predictions, for every class.

ROC-AUC Curve: It serves to evaluate the model on its capacity to separate classes, through discrimination by means of probability thresholds.

4 Result and Analysis

Table 1. Training and Validation Accuracy and Loss

	Training		**Validation**	
Model	**Accuracy**	**Loss**	**Accuracy**	**Loss**
LSTM	0.9953	0.0089	0.9507	0.2156
Bi-LSTM	0.9947	0.0146	0.9412	0.2658
LSTM+Attention	0.9941	0.0138	0.9538	0.3281
Bi-LSTM+Attention	0.9937	0.0210	0.9643	0.1991

From Table.1, one can observe training and validation performances wrt. accuracy and loss for the four models of LSTM, Bi-LSTM, LSTM with Attention, and Bi-LSTM with Attention. On these data, all models onboarded high accuracy, ranging from 99.37% to 99.53%, emphasizing their strong learning ability. The observed difference becomes significant when we come to validation performance: it is the true measure of the generalizability of some model.

The training and validation curves displayed in Fig.4 provide visual support to the tabled findings.

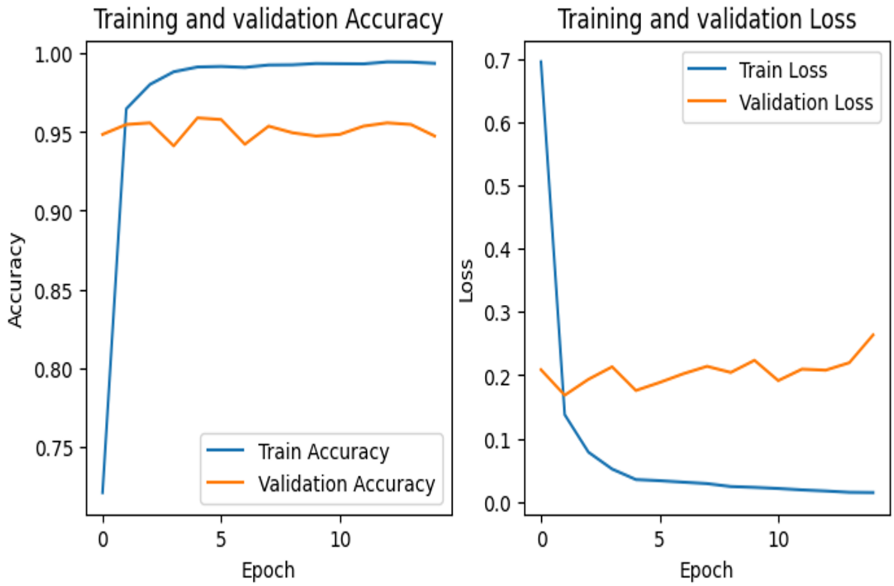

Fig. 4. Accuracy & Loss of Training & Validation for the Best performed Model(Bi-LSTM+Attention)

The Bi-LSTM with attention, exhibiting the best trade-off between training and validation, reached higher accuracy at an even lower validation loss. Both loss and accuracy curves of the model give testimony to smooth convergence in this case. Hence, if taken all together, it would be the best model among the four for Bangla news classification in this study.

Table 2. Comparison of Confusion Metrics for All Models

Model	Accuracy	Precision	Recall	F1-Score
LSTM	94%	0.94	0.94	0.94
Bi-LSTM	95%	0.94	0.94	0.94
LSTM + Attention	95%	0.95	0.95	0.95
Bi-LSTM + Attention	96%	0.96	0.96	0.96

Table.2 shows the metric performances with Accuracy, Precision, Recall, and F1-Score of the four models used in the analysis. The results indicate a consistent improvement as the architectural complexity increases.Unsurprisingly, this is the best-performing model, with an accuracy rate of 96%, as well as precision, recall, and F1-score rates. Hence, this also means that the model is very much accurate, and it is equally balanced in identifying all classes correctly and without discrimination.

Due to the advantage of attention mechanisms largely in boosting contextual representation, the LSTM + Attention model obtains better metrics across all measures versus Bi-LSTM, both giving an accuracy score of 95%. The bare LSTM model is slightly outperformed by 94% accuracy but remains fairly reliable on classification with balanced precision and recall.

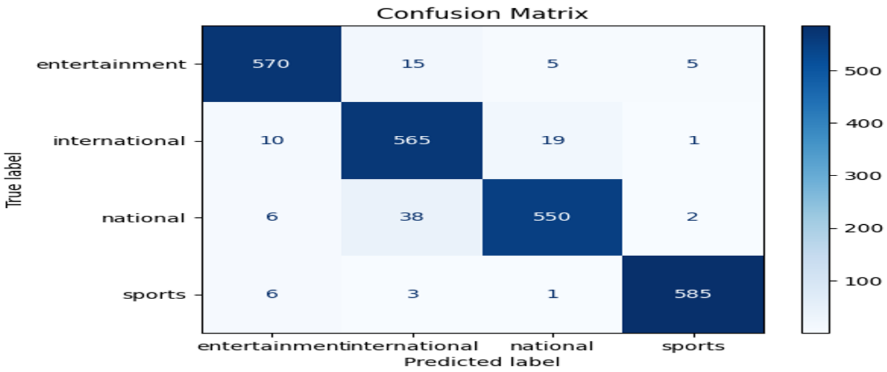

Fig. 5. Confusion matrix of the Best performed Model(Bi-LSTM+Attention)

Table 3. Classification Report of Bi-LSTM+Attention

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Entertainment	0.96	0.96	0.96	595
International	0.92	0.95	0.94	595
National	0.96	0.93	0.94	596
Sports	0.99	0.98	0.98	595
Accuracy	-	-	**0.96**	**2381**
Macro Avg	0.96	0.96	0.96	2381
Weighted Avg	0.96	0.96	0.96	2381

Table 3 presents class-wise evaluation metrics for the last-built model: precision, recall, F1-score, and support. The model shows consistently high performances over all these categories. In classifying Sports, the precision stands at 0.99 and the F1-score at 0.98, which means the model is almost fully confident and highly accurate in discriminating sports news articles. Entertainment comes next with strong performances in all metrics-0.96, which also means this is a very reliable model in distinguishing entertainment news with little misclassification. The National and International classes are almost equally good, with F1-scores for both standing at 0.94. Nonetheless, the recall for the National class is somewhat low at 0.93.

The model convinces itself to generalize beyond training data, building a good impression of itself with respect to the 2,381 test samples exhibiting an overall accuracy of 96%. Using a macro average and weighted average, the value stood at 0.96 for precision, recall, and F1-score. This shows balanced behavior from the classifier in respect to each class, hence making the overall system potent and robust in the presence of slightly varying class distributions. The closer the macro and weighted averages, the less domination of one particular class in model prediction. Thus, there is still a theoretical basis for using the proposed model to classify real Bangla news requiring equal relevance to different topics.

Table 4. AUC Scores of All Models Across News Categories

Model	**Entertainment**	**International**	**National**	**Sports**
LSTM	0.96	0.94	0.95	0.97
Bi-LSTM	0.97	0.95	0.97	0.99
LSTM + Attention	0.98	0.96	0.98	0.99
Bi-LSTM + Attention	**0.99**	**0.98**	**0.99**	**1.00**

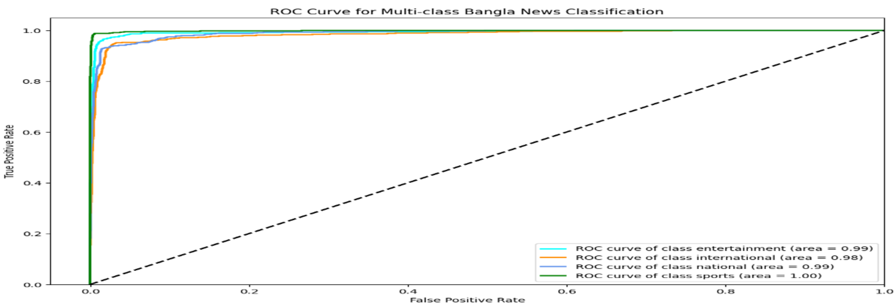

Fig. 6. ROC-AUC Curve for Bi-LSTM+Attention

Further moving towards assessing the models' performance over multiple classes, we trained an ROC curve for each category: entertainment, international, national, and sports. The ROC graphs, plotted in Fig.6 , give the visual representation of the trade-off between the true positive rate and false positive rate encountered for each class. The AUC values were exceptionally high: 0.99 for entertainment, 0.98 for international, 0.99 for national, and a perfect 1.00 for sports. This meant that the models, particularly Bi-LSTM + Attention, are quite successful at distinguishing instances into the four categories. These nearly perfect AUC scores hence imply that the model generalizes quite well and classifies across all classes, hence accepts the validation of proposed attention-augmented architectures.

5 Conclusion

The evaluation of four deep learning models for Bangla news classification was conducted on a Kaggle dataset, with comparatively balanced data, revealing that attention mechanisms greatly improve the performance of LSTM and Bi-LSTM scheme-based models. Among the tested models, Bi-LSTM with Attention stands par excellence, attaining accuracy above 96%, proving the essence of context-aware and semantic understandings for Bangla text processing. These results corroborate the extensive findings that attention-based and transformer architectures are most suited for Bangla NLP tasks that must capture subtle semantic cues. Thus, the future scope of this work lies in the enhancement of classification results using advanced pre-trained language models such as BanglaBERT or multilingual BERT for rich semantic embeddings, exploring zero-shot learning to classify unseen categories, and overcoming the challenges posed by code-mixed Bangla-English data, with the present-day escalation of the phenomenon in digital communication. Hence, this research truly establishes a strong foundation for the development and deployment of a reliable, real-time Bangla news classification system, with wonderful opportunities in media monitoring, social media content filtering, and personalized content recommendations contributing toward the advancement of Bangla language technology.

References

1. M. K. Hasan et al., "Classifying Bengali Newspaper Headlines With Advanced Deep Learning Models," Asian Journal of Research in Computer Science, vol. 16, no. 4, pp. 372–388, 2023.
2. A. S. M. Siam et al., "Bangla News Classification Employing Deep Learning," Communications in Computer and Information Science (CCIS), vol. 1949, 2023.
3. M. K. Hasan and R. Ahmed, "Performance Comparison of Traditional Classifiers for Bangla News Categorization," International Journal of Computer Applications, 2021.
4. U. Roy et al., "Enhancing Bangla News Classification Using Bi-LSTM," International Journal of Advanced Computer Science and Applications, vol. 12, no. 6, pp. 124–130, 2021.

5. M. Z. Hossain et al., “Potrika: Raw and Balanced Newspaper Datasets in the Bangla Language,” arXiv preprint arXiv:2210.09389, 2022.
6. M. Rahman and S. Nahar, “Hybrid CNN-BiLSTM for Bangla News Category Classification,” in Proc. Int. Conf. Computer and Information Technology (ICCIT), 2022.
7. A. Hossain et al., “Potrika: A Large-Scale Dataset for Bangla News Classification,” arXiv preprint arXiv:2210.09389, 2022.
8. M. A. Islam et al., “A Comparative Study of Deep Recurrent Models for Bangla News Classification,” Journal of Intelligent Systems, vol. 32, no. 5, pp. 541–554, 2023.
9. K. Rahman et al., “Text Classification in Bangla Using Attention-Based Deep Networks,” in Proc. IEEE TENCON, 2023.
10. F. Nahar et al., “Improving Bangla News Categorization Through Custom Word Embeddings,” in Proc. Int. Conf. Innovation in Engineering and Technology (ICISET), 2023.
11. N. Sultana et al., “Bangla Text Categorization Using Bidirectional Recurrent Networks,” in Proc. Int. Conf. Computing, Communication and Networking Technologies (ICCCNT), 2022.
12. S. Khan and F. Ahmed, “Comparative Evaluation of Attention Mechanisms for Bangla News Classification,” in Proc. IEEE Int. Conf. Machine Learning and Applications (ICMLA), 2023.
13. M. J. Uddin et al., “Hierarchical Attention Networks for Bangla Document Classification,” in Proc. IEEE Int. Conf. Big Data, 2023.
14. S. Ahmed et al., “Transformer-Based Bangla News Classification: A Case Study,” in Proc. IEEE Int. Conf. Information and Communication Technology (ICICT), 2023.
15. P. Talukder et al., “Benchmarking Bangla News Headline Classification Using Deep Learning,” in Proc. IEEE Int. Conf. Consumer Electronics (ICCE), 2022.
16. M. R. Rashid et al., “Bangla News Summarization and Classification Using Attention Networks,” arXiv preprint arXiv:2305.11522, 2023.
17. S. Dey et al., “A Deep Learning Approach for Classifying Bangla News Headlines,” Procedia Computer Science, vol. 132, pp. 956–963, 2018.
18. T. Rahman et al., “Bangla News Classification with Word Embeddings and LSTM,” in Proc. Int. Conf. Computer and Information Technology (ICCIT), 2022.
19. N. Ahmed and M. Hossain, “Hybrid Deep Models for Bangla Text Categorization,” in Proc. IEEE Int. Conf. Information and Communication Technology (ICICT), 2023.
20. K. Rahman et al., “An Efficient Approach to Categorizing Bangla News Articles Using BiLSTM with Attention,” in Proc. IEEE Int. Conf. Innovation in Engineering and Technology (ICISET), 2023.