Asian Journal of Research in Computer Science

Volume 16, Issue 4, Page 372-388, 2023; Article no.AJRCOS.110372
ISSN: 2581-8260

Classifying Bengali Newspaper Headlines with Advanced Deep Learning Models: LSTM, Bi-LSTM, and Bi-GRU Approaches

Mohammad Kamrul Hasan [a*], Shoeb Akibul Islam [a], Md. Sabbir Ejaz [a], Md. Mahbubul Alam [a], Nahid Mahmud [a] and Tanvir Ahmed Rafin [a]

[a] Department of Information and Communication Engineering, Noakhali Science and Technology University, Bangladesh.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJRCOS/2023/v16i4398

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/110372

Original Research Article

Received: 10/10/2023
Accepted: 13/12/2023
Published: 16/12/2023

ABSTRACT

Reading newspapers is beneficial for people of all ages and the global community. The enjoyment of gathering diverse data from various sources adds to the overall experience. To enhance specificity in Bengali news headlines, recognizing the news genre becomes crucial. Recognizing the genre of the news, it is a very challenging task in Bengali Text Classification with the help of AI. A very few research works is done on Bengali News headline classification and we have done a model to provide a solution to the addressed issue. Due to the continuous change of the structure of the news headlines, we have employed a neural network adoption connection to our methodology

**Corresponding author: E-mail: Kamrul.ice@nstu.edu.bd;*

experiment on a mixture of primary and secondary dataset. Achieving significant results, we implemented a Bengali dataset in Multi Classification using Long-Short Term Memory (LSTM), Bi-Long-Short Term Memory (Bi-LSTM), and Bi-Gated Recurrent Unit (Bi-GRU). The dataset is established by aggregating news headlines from various Bengali news portals and websites, showcasing robust categorization performance in the end product. Six categories were employed for the classification of Bengali newspaper headlines. The Bi-LSTM Model emerged with the highest training accuracy at 97.96% and the lowest validation accuracy at 77.91%. Furthermore, it demonstrated enhanced sensitivity and specificity.

Keywords: Bengali news; news classification; machine learning; deep learning; social newspaper; news headlines.

1. INTRODUCTION

Structured formatted data, such as news data, encompasses text-based attributes like source, date, location, author, headline text, and detailed information. Language processing approaches are commonly favored for extracting features from textual data, as they effectively tackle various linguistic and text-related challenges [1].

Text classification remains a dynamic area of research, with a growing interest among researchers in developing text mining applications that leverage classification techniques. It's worth noting that a significant portion of existing text classification algorithms has been predominantly evaluated on datasets composed in the English language. Research on Bengali text classification is currently limited. Bengali, being one of the most extensively spoken languages globally, has become a focal point in the realms of natural language processing (NLP) and text classification. The growing prevalence of digital content in Bengali has underscored the need for the creation of robust techniques aimed at organizing, analyzing, and extracting insights from textual data in the Bengali language. Machine learning and deep learning methodologies have emerged as potent tools for Bengali text classification, facilitating the automated categorization of Bengali documents into predefined classes [2]. Classifying Bengali text poses a challenge due to the morphologically rich nature of the language. The feature vector dimension tends to be exceptionally high, particularly with large corpora, adding to the complexity of the task.

The aim of this investigation is to explore the challenge of classifying Bengali newspaper headlines through the utilization of deep learning models, specifically LSTM, Bi-LSTM, and Bi-GRU models [3]. These models are widely used for sequence modeling and have shown promising results in a variety of natural language processing (NLP) applications, especially in text categorization. In order to accomplish this, we collected a dataset of headlines from Bengali newspapers and preprocessed the information in order to get it ready for model training. We then used this dataset to train and assess the LSTM, Bi-LSTM, and Bi-GRU models. To improve their performance, we used a variety of strategies, including regularization and data augmentation. The proposed methodology entails gathering and preprocessing a dataset of news articles written in Bengali, using trained word embeddings to represent text, building and training multiple neural network architectures, experimenting with hyperparameters, evaluating performance with metrics, testing the best-performing model, and thoroughly analyzing the outcomes. The importance of precise text classification in news headlines cannot be overstated. Given the prevalence of fake news and disinformation, the ability to differentiate between trustworthy and untrustworthy sources of information is critical. Furthermore, effective text classification aids in information organization and retrieval, ultimately improving the user experience.

The primary objective of this study is to evaluate the effectiveness of the proposed methodology in accurately categorizing Bengali news headlines. The study's findings hold relevance across various domains, including content filtering, information retrieval, and topic modeling, contributing to the development of language-specific text categorization systems. In this research, we delve into the utilization of deep learning models, specifically LSTM, Bi-LSTM, and Bi-GRU models, for the classification of Bengali newspaper headlines to address these challenges. While not extensively explored in the realm of Bengali text categorization, these models have shown promise in other NLP applications such as sentiment analysis and named entity recognition.

The subsequent sections of the paper follow this structure: Section 2 provides an overview of previous studies, while Section 3 outlines our proposed approach. To assess the accuracy of our suggested approach and determine the most effective deep learning method, Section 4 conducts an experimental analysis of each algorithm. Finally, Section 5 concludes the paper by summarizing its key points and suggesting future directions.

2. LITERATURE REVIEW

Throughout the years, extensive research has been conducted in the field of headline detection. Beyond the realm of classifying Bengali headlines using textual data, there exist numerous other areas worthy of exploration. This section will delve into previous research endeavors focused on the classification of Bengali newspaper headlines based on textual data.

Researchers of article [4] employed CNNs to categorize Bengali text into five categories in a study titled "Bengali Text Classification using Convolutional Neural Networks with Fast Text Embedding," attaining an accuracy of 96.85%. In essence, this publication offers a comprehensive overview of previous research efforts in Bengali text classification.

The review work presented in [5] covers the text classification process, classifiers, and various feature extraction approaches, with a specific focus on short texts, exemplified by the classification of news articles based on their headlines. The discussion encompasses machine learning algorithms such as K-Nearest Neighbors, Naive Bayes, Support Vector Machines, Artificial Neural Networks, and Decision Trees.

Hossain and Chaudhury in "Bengali News Headline Categorization," extracted news headlines and their categories from a number of online publications [6]. This study considers eight news categories, classifying news based on their headlines. The input data is modeled using LSTM and GRU neural networks, and the predicted category is compared with the actual category. The GRU model achieves an accuracy of 87.48%, outperforming the LSTM model, which attains an accuracy of 82.74%. In terms of accuracy, GRU surpasses LSTM in this context.

According to the study by Khushbu and Masum et al., specifying the news genre in Bengali news headlines can enhance their specificity to a greater extent [7]. Consequently, the machine can intelligently analyze the sequence of sentences in the output to discern the news genre. Through experimentation, the authors establish a correlation between the adoption of our strategy and a 90% accuracy rate. The authors conducted Multi Classification on a Bengali dataset using SVM, NB, Logistic Regression, Neural Network, and Random Forest, yielding notable outcomes.

A paper published by Gurmeet Kaur & Karan Bajaj named “News Classification and Its Techniques: A Review”, in their paper they have considered the problem of classification off news classification [8]. They have presented an algorithm for category identification of news and have identified number of shortcomings of a number of algorithms approaches. Thorough exploration is conducted into processes such as pre-processing, document indexing, feature selection, and news headline classification. Furthermore, these algorithms can be refined to enhance the efficiency of categorization.

In the study titled "n-Bi-LSTM: Bi-LSTM with n-gram Features for Text Classification," researchers introduce a method to develop a precise and swift text classification system, employing both One-vs.-one and One-vs.-rest approaches. Their approach, named n-Bi-LSTM, transforms natural text sentences into features resembling bag-of-words with n-gram techniques, subsequently feeding these features into a bidirectional LSTM. The results demonstrate the superiority of the n-Bi-LSTM algorithm over basic LSTM and bidirectional LSTM algorithms [9].

The paper titled "Real Time News Classification Using Machine Learning" [10] focuses on leveraging headlines to categorize real-time news. A systematic categorization assigns each news headline to predefined categories using a dedicated system. To enhance accuracy beyond standalone methods, a hybrid model leveraging various algorithms was developed. This classifier operates in real-time, retrieving and processing news headlines. Through a comparative examination of multiple classifiers, this approach not only yields a more effective working model but also demonstrates superior true positive rates. The fusion of SVM and LR classifiers resulted in a hybrid model, surpassing SVM and LR models by 0.13% and 0.16%, respectively. Trained on the same dataset, the hybrid model achieved an accuracy of 89.79%. This hybrid model is employed for real-time data retrieval and categorization based on headline content.

In the study titled "Sentiment Analysis of Bengali News using Deep Learning Methods," scholars employed Long Short-Term Memory (LSTM) and Bidirectional LSTM (BI-LSTM) techniques to assess sentiment. The obtained accuracy rates were 80.52% for LSTM and 81.78% for BI-LSTM [11].

In the research conducted under the title "Sentiment Analysis of Bengali Newspaper Headlines," the LSTM model was utilized. The findings revealed an accuracy of 83.12% in employing LSTM for the sentiment analysis of Bengali newspaper headlines. The study underscores the superior accuracy of the LSTM model, positioning it as the most effective approach. [12].

In the paper titled "Classification of Bengali Text Using Deep Learning Techniques," researchers utilized BI-LSTM, Bi-GRU, and CNN to categorize Bengali text into seven distinct classes. The obtained accuracy rates for these models were 91.57%, 92.22%, and 90.63%, respectively. Notably, BI-LSTM, Bi-GRU, and CNN emerged as the top-performing models, demonstrating the highest accuracy in the classification task. [13].

Using ongoing electroencephalogram (EEG) inputs, the authors of article [14]_present a hierarchical bidirectional Gated Recurrent Unit (GRU) network for categorizing human emotions. The architecture of the model mirrors the hierarchical nature of EEG signals, incorporating an attention mechanism at two levels—EEG samples and epochs. This dual-level attention allows the model to discern varying levels of information relevance, enabling the acquisition of a more pertinent feature representation of the EEG sequence. This approach accentuates the importance of specific samples and epochs in contributing to the model's affective category predictions. To assess the model's performance, the authors conducted cross-subject emotion categorization tests using the DEAP dataset. The experimental findings reveal that, in both valence and arousal dimensions, their model, applied to 1-second segmented EEG sequences, surpasses the performance of the best deep baseline LSTM model by 4.2% and 4.6%, respectively. Additionally, compared to the best shallow baseline model, their model exhibits improvements of 11.7% and 12%. Furthermore, these advantages persist as the epoch length increases.

In the research titled "Bengali News Classification using Deep Learning Techniques: A Comparative Study," the authors evaluated the effectiveness of CNN, LSTM, BI-LSTM, and GRU in the classification of Bengali news articles. The results indicated that BI-LSTM outperformed the other models, achieving the highest accuracy at 93.34% [15]. Outperforming baseline methods in the classification of EEG sequences, this model proves its effectiveness in addressing the challenges associated with the extended non-stationarity of EEG data. As a result, the proposed model exhibits enhanced accuracy and robustness in the field of emotion classification based on EEG signals.

In the study titled "Bengali News Classification Using Deep Learning Techniques," CNN, LSTM, and BI-LSTM were employed by researchers to classify Bengali news into six categories, achieving accuracy rates of 89.63%, 89.20%, and 90.22%, respectively [16].

Conducted by G. Bharath, K. J. Manikanta, and G. Bhanu Prakash, the investigation documented in [17] delves into the significance of online media as an information source. Concentrating on identifying fake news disseminated on Twitter, the authors advocate for a model that incorporates machine learning algorithms, specifically Support Vector Machine and Naïve Bayes. The experimental outcomes underscore the effectiveness of these approaches in automated fake news detection, underscoring the pivotal role of automation in upholding the credibility of online media and social networks.

In a paper titled "Deep Learning-Based Text Classification for Bengali News," researchers proposed a new way of classifying Bangla news documents using a Deep Recurrent Neural Network. The Deep Recurrent Neural Network with Bi-LSTM achieved 98.33% accuracy which is higher than other well-known classification algorithms in Bangla text classification [18].

Among all the above all the research articles, the accuracy is comparatively quite low and it needs to be improved. Our research work is also focusing on primary data because of the continuous change of news pattern. The ultimate aim of this research work is to improve the classification accuracy of Bengali news headline and check the how well the proposed model performed with the acquired dataset and the deep learning models.

3. METHODOLOGY

This study aims to classify Bengali newspaper headlines, employing Natural Language Processing (NLP) methods and extracting features from deep learning models to preprocess the Bengali headlines. Subsequently, we will assess the performance of the models. Various Python libraries, including scikit-learn, TensorFlow, NLTK, and Matplotlib, will be utilized throughout the project for implementation and analysis. Fig. 1 illustrates the proposed methodology for this study. The suggested approach utilizes deep learning models and encompasses several steps aimed at the precise categorization of Bengali news headlines. The methodology encompasses processes such as data collection, preprocessing, word embeddings, the suggested model, model evaluation, and more.

3.1 Dataset Collection

In response to the evolving data patterns, we opted to collect raw data and apply machine learning methods for analysis. The data collection process took two distinct approaches. A portion of the data was sourced from publicly available datasets online, while the remaining dataset was obtained in its raw form from various Bengali newspapers in the social domain. Several web sources, including “Prothom Alo” [19], “Daily Ittefaq” [20], “Bangladesh Pratidin” [21], “Kaler kontho” [22], “Daily Inqilab” [23], and “Jugantor” [24] were used to gather the data that we have used. Better outcomes are anticipated with increased accessibility to data. In the context of our news headline classification task, a substantial volume of headlines was deemed essential for optimal results. Illustrated in Fig. 2 is a representation of raw data extracted from two online newspapers. Our dataset for the investigation comprises 1000 primary data and 9000 secondary data, resulting in a total of 9913 samples after the cleaning process. The dataset is characterized by six class tags, facilitating the categorization of news items into distinct groups such as politics, national interest, entertainment, sports, international, and IT. The frequency distribution of headlines within each category is outlined in Table 1. This dataset is identified as "Bengali Newspaper Headline".

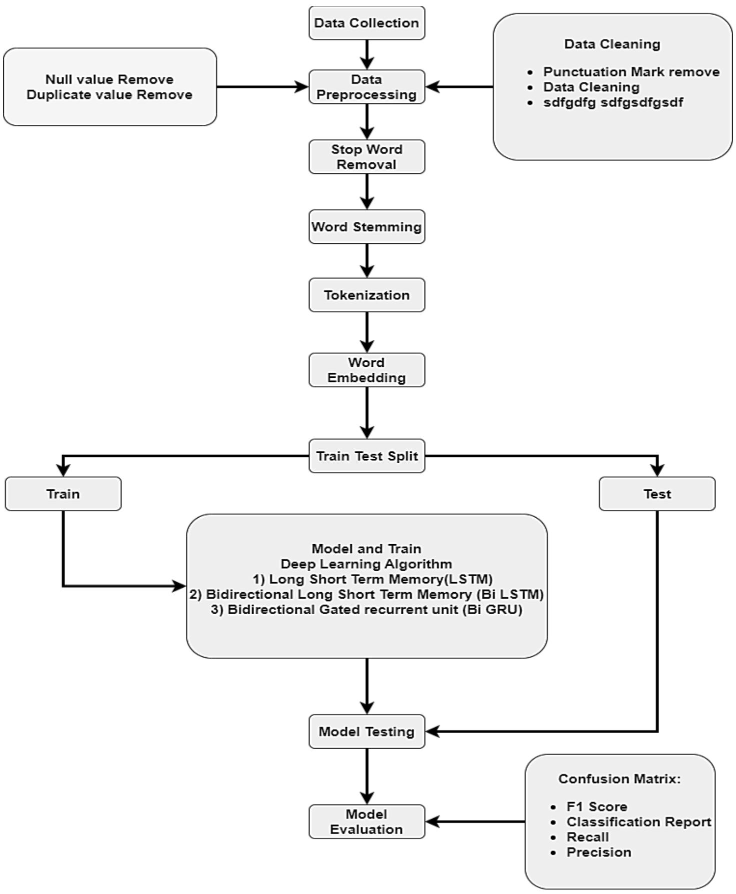


Fig. 1. Proposed model for the classification of Bengali Newspaper headlines

Fig. 2. Some Headline of Bengali Newspaper

Table 1. Category of Dataset with a count

Class	Count
Politics	1692
National	1689
Amusement	1673
Sports	1630
International	1628
IT	1601

3.2 Data Preprocessing

Data preprocessing is a crucial technique for converting unstructured data into a format that can be easily interpreted. This stage is of paramount importance because real-world data often contains flaws, inconsistencies, missing values, or lacks key attributes. The primary objective of data preprocessing is to clean and organize data, making it more accessible for interpretation by machine learning models. Failure to preprocess data can result in the presence of numerous null values, leading to ambiguity in machine learning and deep learning processes. In our dataset, several preprocessing techniques have been employed. Elimination of punctuation marks, handling NaN values, and removing duplicate entries are integral steps in the data preprocessing pipeline. Punctuation marks, while serving as symbols to convey the structure and meaning of sentences, are often disregarded in the analysis of text data. However, their removal contributes to refining the dataset and enhancing the clarity of machine learning and deep learning procedures. Bengali punctuation examples include the following "।", "একটি বিন্দু", and "বাংলা এমনকি". Punctuation has been removed, and the outcome is seen in Fig. 3. Fortunately, our collection doesn't contain any Nan values.

3.3 Stop Words Removal

We have used an additional preprocessing method to remove the stop words. Stop words are terms that are frequently used in a language but have no significant meaning when it comes to text analysis. Examples of these words are headings, conjunctions, and prepositions. We have employed Natural Language Toolkit to remove the stop words from the headlines. An example of the list of stop words is shown in Fig. 4.

	headline	category	newspaper name	After Preprocessing
0	'১ মিনিটে নগদ অ্যাকাউন্ট' সেবা উদ্বোধনে জয়	IT	Jugantor	মিনিটে নগদ অ্যাকাউন্ট সেবা উদ্বোধনে জয়
1	'১৫ হাজার কোটি টাকা ব্যয়ে সৈয়দপুর বিমানবন্দরের...	national	Dainik Ittefaq	হাজার কোটি টাকা ব্যয়ে সৈয়দপুর বিমানবন্দরের কা...
2	'২ ঘণ্টায় সালমানের গ্যালাক্সি অ্যাপার্টমেন্ট ব...	Amusement	Jugantor	ঘণ্টায় সালমানের গ্যালাক্সি অ্যাপার্টমেন্ট বোম...
3	'২ বছর নিষিদ্ধ হতে পারতেন রোনাল্ডো'	sports	Jugantor	বছর নিষিদ্ধ হতে পারতেন রোনাল্ডো
4	'৩০৮ জন নারীর সঙ্গে শারীরিক সম্পর্ক সঞ্জয় দত্তের'	Amusement	Jugantor	জন নারীর সঙ্গে শারীরিক সম্পর্ক সঞ্জয় দত্তের

Fig. 3. Before and after Removing Punctuation on Headline

6	4	অথচ
7	5	অথবা
8	6	অধিক
9	7	অধীনে
10	8	অধ্যায়
11	9	অনুগ্রহ
12	10	অনুভূত
13	11	অনুযায়ী
14	12	অনুরূপ
15	13	অনুসন্ধান

Fig. 4. Some stop words from our dataset

3.4 Word Clouding

A word cloud is a visual representation of text data where each word's magnitude indicates how frequently or important it occurs. This popular technique for displaying text data can be used to rapidly identify the terms that occur most frequently in a text. The “WordCloud” library is used to peroform this task. Figs. 5 and 6 display an example of the word clouds for the national and political categories.

3.5 Word Stemming

Word stemming reduces a word to its most basic form. For example, "Working", "Work", and "Worker" will all become "Work" after stemming. For rapid and accurate classification, stemming is necessary. Reducing words in Bengali to their lemma, or base or root form, is known as Bengali stemming. During this process, the affixes (suffixes and prefixes) are removed from the words in order to classify various spellings of the same term into a single category. For this reason, we have stemmed the Bengali words using the "Bangla-stemmer" library. Fig. 7 displays the output of the headlines following the application of the stemmer.

3.6 Label Category

The LabelEncoder() function can be used to translate category variables into numerical values. It assigns a unique numerical value to every category in the provided data, which may be fed into machine learning models that require numerical input. When using LabelEncoder() to encode categories, it is important to remember that the encoding process is random and does not transmit any inherent ranking or order among the categories. Put otherwise, the numerical values assigned to the categories do not imply any meaningful link between them. For example, the numerical values (e.g., 0, 1, 2) assigned to newspaper categories such as "politics," "sports," and "amusements" do not indicate that one category is more important or ranked higher than the others. Fig. 8 provides a sample for marking the categories. Figs. 9 and 10 show the labeling of the categories in our dataset.

Fig. 5. Word cloud of politics

Fig. 6 Word cloud of National

মিনিটে নগদ অ্যাকাউন্ট সেবা উদ্বোধনে জয়
টাকা ব্যয় সৈয়দপুর বিমানবন্দর শিগগির
ঘণ্টায় সালমান গ্যালাক্সি অ্যাপার্টমেন্ট বোমা ম...
নিষিদ্ধ পারতেন রোনাল্ডো
নারীর শারীরিক সম্পর্ক সঞ্জয় দত্ত

Fig. 7. After Applying Bengali stemmer.

```
Category         label
-------------   -------
Amusement             0
IT                    1
International         2
national              3
politics              4
sports                5
```

Fig. 8. Category label

```
politics
politics
politics
Amusement
IT
IT
IT
national
IT
sports
```

Fig. 9. Before labeling 1st 10 Category

```
4
4
4
0
1
1
1
3
1
5
```

Fig. 10. After labeling1st 10 Category

3.7 Word Embedding and Tokenization

A text is divided into tokens, which are smaller units, using this method [25]. Words, letters, or sub words can all be used as these tokens. For instance: "১০ মার্চ থেকে প্রতিযোগিতা শুরু!", Tokenized into the following line like "১০" "মার্চ" "থেকে" "প্রতিযোগিতা" "শুরু" "!" [26]. Moreover, word embedding converts words to their vector or numerical values. The "One Hot" approach was employed for both processes.

A) One hot: We used one-hot encoding to encode the Bengali newspaper headlines as numerical data for classification in this study [27].

মিনিটে নগদ অ্যাকাউন্ট সেবা উদ্বোধনে জয়
টাকা ব্যয় সৈয়দপুর বিমানবন্দর শিগগির
ঘণ্টায় সালমান গ্যালাক্সি অ্যাপার্টমেন্ট বোমা মার উড়া দেয়া
নিষিদ্ধ পারতেন রোনাল্ডো
নারীর শারীরিক সম্পর্ক সঞ্জয় দত্ত
বছরে ইরানকে কাবু পারেনি যুক্তরাষ্ট্র
বন্দি খালেদা জিয়া চিরপঙ্গু দেয় ষড়যন্ত্র চল
অগমেডিক্স বাংলাদেশ অগমেডিক্স পায় টাকা
অত্যাচারের মুখে বিএনপি শক্তিশালী
অব লুটেরাস বাই লুটেরাস ফর লুটেরাস

Fig. 11. Before one Hot Tokenizer in 1st 10 headlines

Using the one-hot encoding technique, numerical data can be represented as categorical data. With the exception of the element that corresponds to the category, every element in a binary vector representing a category has a length equal to the total number of categories. A snapshot of the dataset before and after tokenization is shown in Figs. 11 and 12.

B) Padding: After the one-hot process, each sequence is padded into a predefined length. Post padding has been utilized in the sequences in our work. An example of a padded sequence for a length of 15 is shown in fig. 13 below.

```
[2556, 372, 2028, 1431, 507, 873],
[1578, 663, 1248, 459, 375],
[1945, 191, 579, 1403, 1211, 842, 899, 1712]
[2374, 1152, 412],
[2919, 83, 634, 639, 2333],
[413, 2082, 1595, 827, 574],
[2021, 1482, 2637, 1376, 1595, 1454, 2966],
[2327, 1164, 2327, 361, 1578],
[2157, 2171, 906, 1183],
[1185, 570, 2317, 570, 2570, 570],
```

Fig.12. After one Hot Tokenizer in 1st 10 headlines

```
array([[ 826, 1883,   16, 1978, 1718,  831,    0,    0,    0,    0,    0,
           0,    0,    0,    0],
       [2601, 1379,  434, 1078, 1901,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0],
       [1348, 2578, 2558, 2908, 2556,  695, 2766, 1426,    0,    0,    0,
           0,    0,    0,    0],
       [ 362,  469, 1088,    0,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0],
       [1639, 2577, 2026, 2752,  676,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0]], dtype=int32)
```

Fig. 13. A padding sequence of length 15

3.8 Train and Test Data Splitting

Following preprocessing and feature selection, the datasets are divided into train and test datasets with 90% and 10% of each, respectively. The training sets will be used to train the models. Testing sets will be used to evaluate the models.

3.9 Deep Learning Models

The proposed model is based on the input parameters of the dataset. After preprocessing, the dataset is fed into recommended deep learning models. For this classification, we used three models to forecast news headlines, which are as follows:

- Long-short-term memory (LSTM)
- Bidirectional Long-short-term memory (Bi-LSTM)
- Bidirectional Gated Recurrent Unit (Bi-GRU)

3.9.1 Long short-term memory (LSTM)

Recurrent neural networks (RNNs) with Long Short-Term Memory (LSTM) are specifically made to process sequential input, such voice, time series, and other types of data. An RNN's current step receives the output from the preceding phase as input. The long-term RNN dependency problem, wherein the RNN may make more accurate predictions based on current data but cannot forecast the word stored in the LSTM, was addressed by Hochreiter & Jürgen Schmidhuber when they built the LSTM. As the gap duration grows, RNN's performance becomes disorganized. LSTM has the ability to store data for a long period of time by default. It is used for prediction, categorization, and time-series data processing. An LSTM consists of three gates: an input gate (it), an output gate (ht), and a forget gate (ft), which regulates how much data is input from the previous memory state [28].

3.9.2 Bidirectional long short-term memory (Bi-LSTM)

The Bi-LSTM (Bidirectional Long Short-Term Memory) model is an LSTM variant that has the ability to handle input sequences in both forward and backward directions. This allows the model to capture both the past context and the future context of each input letter, which makes it especially suitable for sequence-to-sequence tasks like speech recognition and natural language processing. Two separate LSTM networks, one processing the input sequence forward and the other backward, make up the architecture of the Bi-LSTM model. Each LSTM network is composed of three gates: the input gate, forget gate, and output gate. The input gate controls information entering the cell state, the forget gate controls information staying in the cell state or deleting it, and the output gate controls information leaving the cell state [29].

3.9.3 Bidirectional gated recurrent unit (Bi-GRU)

Bi-GRU (Bidirectional Gated Recurrent Unit) is a version of the Gated Recurrent Unit (GRU) model that is capable of both forward and backward analysis of input sequences. This allows the model to record the past and future context of each input token, similar to the Bi-LSTM. This makes the model particularly suitable for sequence-to-sequence tasks such as natural language processing and speech recognition. The Bi-GRU model uses GRU units instead of LSTM units, which is the primary architectural difference between the Bi-GRU and Bi-LSTM models. Apart from a cell state, GRU units also have two other gates: the reset gate and the update gate. While the reset gate controls whether prior information is retained or deleted, the update gate controls the flow of new information into the cell state. At each time step, the input sequence is first transmitted into the forward GRU network, producing a number of forward hidden states. The input sequence is also supplied to a reverse GRU network, which generates a number of backward hidden states. A fused representation formed by concatenating the forward and backward hidden states is utilized to encode the past and future context of the input token. This fused representation is then forwarded to a fully linked layer for classification or further processing [30].

3.10 Model Testing

Using our trained LSTM, Bi-LSTM, and Bi-GRU models, we have predicted the class of the newspaper headlines in this phase. In this case, the headlines' placement across a variety of categories—sports, entertainment, health, the world, and politics—is referred to as the predict categorization. Using these news headlines, we forecast the classification and compare the predicted classified class with the original class.

3.11 Performance Parameters

In this study, performance will be compared using several evaluation criteria. It will ascertain how effectively the model is able to distinguish and classify headlines. Examining standard metrics is essential to understanding the performance of competing models. The four elements of the confusion matrix—accuracy, precision, recall, and f1-score—are widely used to assess how well deep learning systems operate.

A) Confusion Matrix: One of the easiest methods for assessing the efficacy and accuracy of the model is to use the confusion matrix [31]. It can assign many classifications to an outcome and is used to address classification difficulties. The two dimensions "Actual class" and "Predicted class" are present in each dimension of the confusion matrix, which is a table. Rows show the actual categories, whereas columns show the predicted classes. Class 0 and Class 1 are the two classes included in the dataset. Additionally, multiclass prediction systems can make advantage of it. The confusion matrix is displayed in Fig. 14 where TP, FP, TN, and FN refer to “True Positive”, “False Positive”, “True Negative”, and “False Negative” respectively.

B) Precision: Positive predictive value, or precision, is a metric used to assess how reliable optimistic forecasts are. This measure is computed as the ratio of true positives to all predicted cases designated as positive. The fraction of true positives among all cases classified as positive is the definition of accuracy in mathematics.

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

C) Recall: The recall of a machine learning algorithm gauges its ability to recollect crucial information. It is computed by: Its primary objective is to identify and collect significant data.

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

D) Accuracy: Accuracy is a critical criterion for assessing an algorithm's efficacy. It is calculated using the following formula, which quantifies how closely analytical results match the actual value:

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

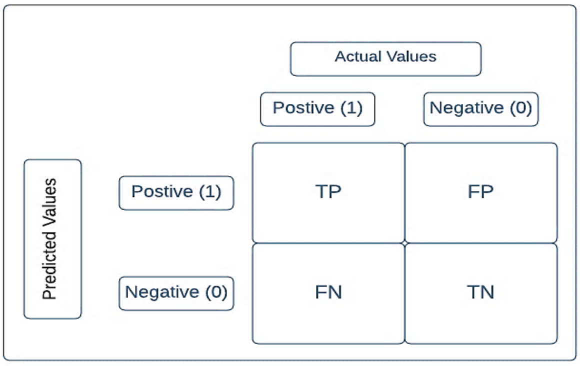


Fig. 14. Confusion Matrix

E) F1 Score: The F1-score is an estimate that accounts for both recall and precision. A person can only obtain an F1 score of 1 of 1 if their recall and precision are perfect. The F1 score is high when recall and precision are both good. When striking a balance between recall and precision is crucial, the F1 score—which is determined by taking the harmonic mean of precision and memory—performs better than accuracy.

$$F1\ Score = 2 * \frac{Precison * Recall}{Precision + Recall}$$

4. RESULT AND DISCUSSION

As previously said, this work uses three Deep Learning techniques to predict the kind of newspaper headline. These techniques' effectiveness has been assessed, and a comparison of them has also been completed. The next part contains the performance evaluation of the different classification methods.

4.1 LSTM

After the model was constructed, we used 40 iterations and the training dataset to train the model. The training accuracy increased to 95.86% in the most recent epoch, however the validation accuracy did not increase at the same rate. Fig. 15's curve shows our validation accuracy for the most recent era, which was 76.29%. Fig. 16 displays the confusion matrix for the LSTM model. It shows that 125 out of 177 data samples had the amusement (class 0) category news accurately categorized. The model mislabeled 52 samples; of those, 12, 12, 11, 2, and 15 samples were incorrectly categorized as belonging to the IT, International, National, Politics, and Sports news categories. Out of 167 samples, the model accurately categorized 104 headlines for the IT category of news, and the largest number of misclassifications occurred for the fun category, with 18 examples. For the International, National, Politics, and Sports categories of news, the corresponding correctly classified samples are 125 (out of 148), 89 (out of 153), 143 (out of 173), and 141 (out of 171). These results provided us with a weighted average that has the following values: 73.23% precision, 73.98% recall, and 73.52% f1 score. The bar chart in Fig. 17 shows the classification report of LSTM model.

4.2 Bi-GRU

Bi-GRU has 20 epochs of completion. The training accuracy increased to 97.96% in the most recent epoch, but the validation accuracy did not increase at the same rate. Figs. 18 and 20 illustrate our 76.10% validation accuracy in the most recent period. Fig. 19 displays the confusion matrix for the Bi-GRU model. The model, which is an improvement over the LSTM model, predicted 131 and 46 correctly and incorrectly, respectively, for Class 0 (amusement). Compared to the previous model, the results of Class 3, 4, and 5 are better since the samples 127, 95, and 147 were correctly classified. Class 1 and Class 5 continue to yield the same results, with 104 and 141 accurate predictions, respectively. Fifty samples are incorrectly categorized as amusement type news during the "Amusement" category classification process. In a similar vein, the samples 34, 57, 41, 23, and 42 are incorrectly categorized as news of the IT, international, national, political, and sporting types. The accuracy of predictions is marginally higher than that of the LSTM model. With an accuracy, recall, and f1 score of 73.23%, 73.98%, and 73.52%, respectively, it produces a weighted average of all classes. The classification reports are displayed in a bar chart of Fig. 20.

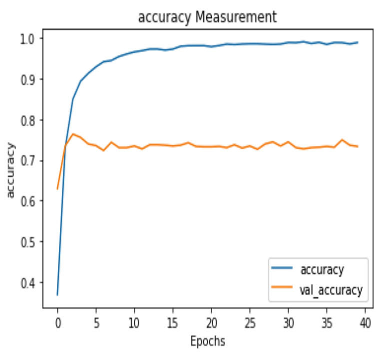


Fig. 15. Training and Validation Accuracy curve for LSTM

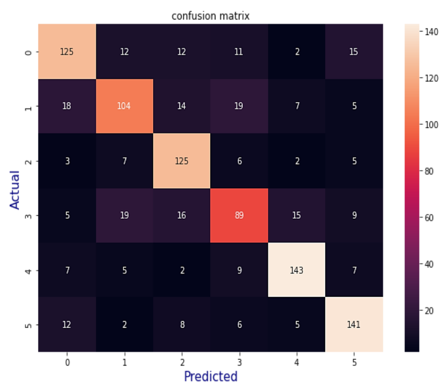


Fig. 16. Confusion Matrix of LSTM

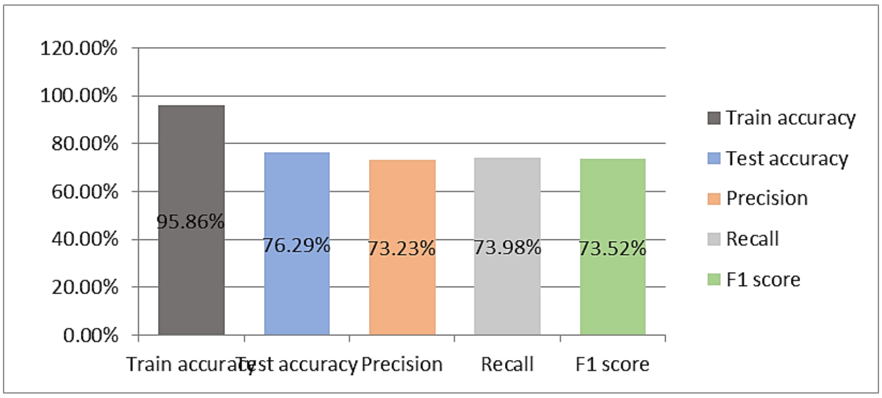
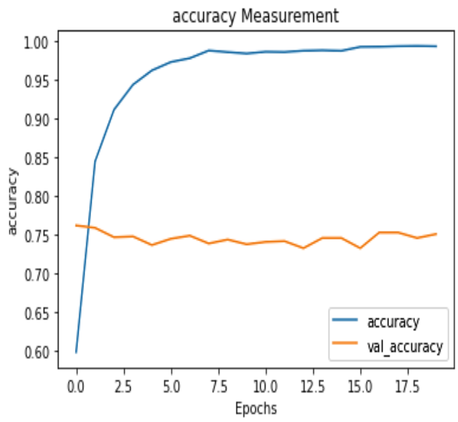


Fig. 17. Chart of Classification Report for LSTM model



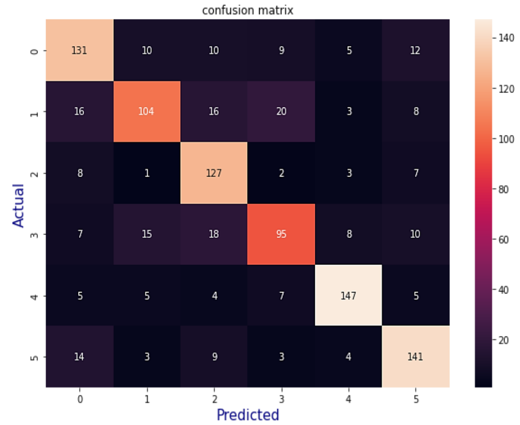


Fig. 18. Confusion Matrix of Bi-GRU model **Fig. 19. Confusion Matrix of Bi-GRU model**

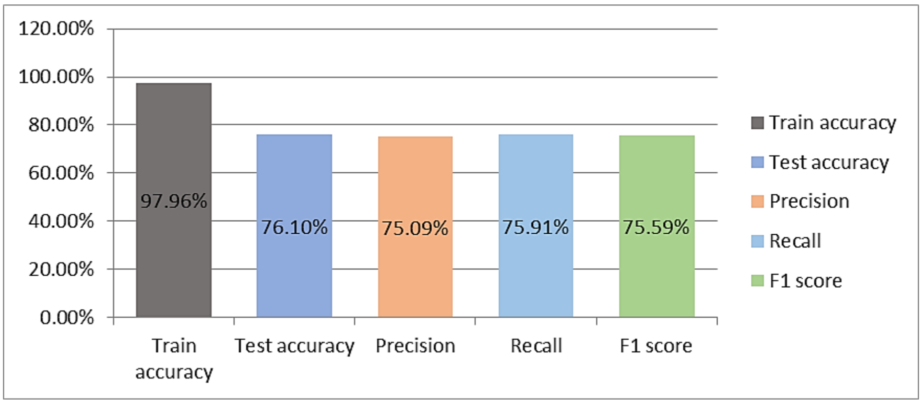


Fig. 20. Chart of Classification Report for Bi-GRU model

4.3 Bi-LSTM

The Bi-LSTM model has been run through a total of 30 epochs in order to ascertain the accuracy of testing and training. The training accuracy increased to 97.96% in the most recent epoch, but the validation accuracy did not increase at the same rate. Our validation accuracy for the prior period was 77.91%. The relationship between training and testing accuracy is shown in Fig. 21. The confusion matrix for the six categories of newspaper headlines is shown in Fig. 22. The algorithm accurately identified 133 samples, mistakenly classified 44 samples, and incorrectly classified 57 samples as belonging to the amusement category (class 0). Additionally, the model has accurately identified 118, 122, 96, 146, and 138 samples for the IT, International, National, Politics, and Sports categories; the false positive samples for these categories are, respectively, 49, 26, 57, 27, and 36. In a similar vein, the Bi-LSTM models also gave inaccurate predictions for 29, 43, 51, 20, and 39 samples, corresponding to IT, International, National, Politics, and Sports news. The matrix yielded a weighted average with 76.15% precision, 76.89% recall, and 76.66% f1 score across all classes. In Fig. 23, all of the matrices are displayed.

4.5 Comparative Analysis

We compare the three deep learning models that were utilized in the study with each other. The performance matrices of each model are compared to determine which produces the best result. Table 2 and Fig. 24 present a comparison of the algorithms' accuracy. From which we can infer that Bi-LSTM has the best test accuracy of 77.91%, while Bi-GRU and Bi-LSTM both have training accuracy of 97.96%. Because just 10% of the data are used, the testing accuracy is comparatively lower.

Fig. 25 and Table 3 present a comparison of the performance metrics of all deep learning models. The results indicate that the precision values of LSTM and Bi-LSTM are quite close to each other, at 76.23% and 76.15%, respectively. When it comes to F1 Score, sensitivity (recall), and overall performance, Bi-LSTM outperforms other models. Finding the optimal algorithm to classify newspaper headlines more correctly than prior approaches is the main goal of this research.

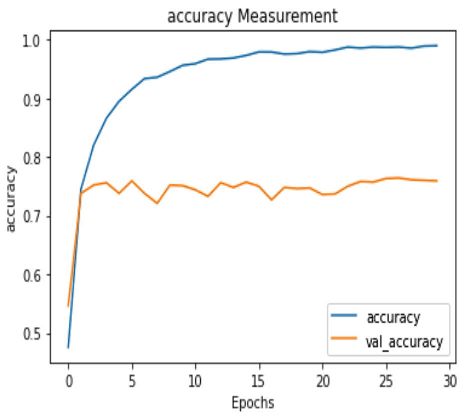


Fig. 21. Training and Validation Accuracy curve for Bi-LSTM

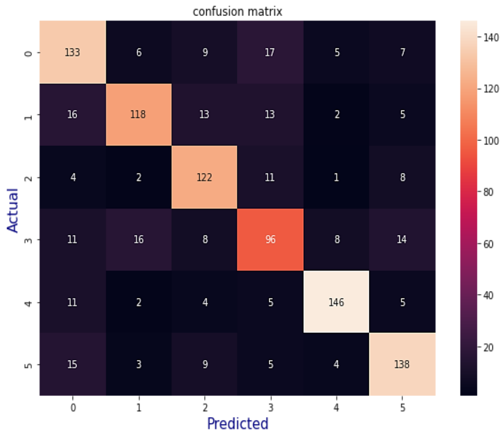


Fig. 22. Confusion Matrix for Bi-LSTM model

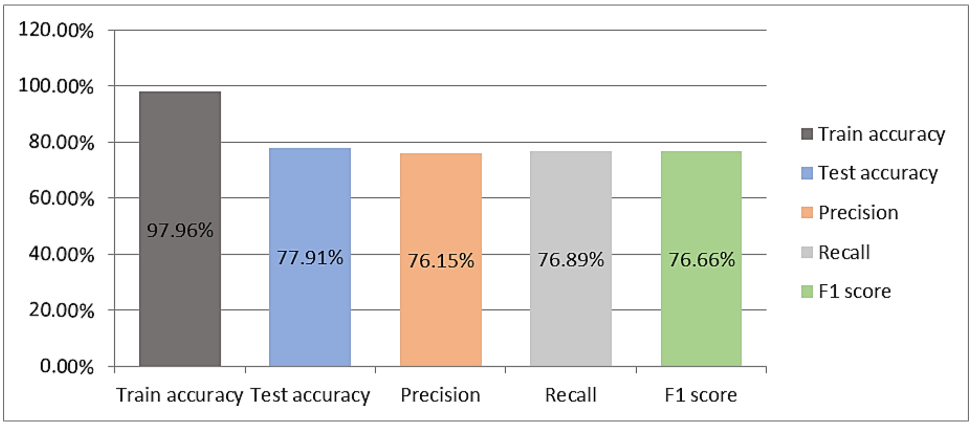


Fig. 23. Chart of Classification Report for Bi-LSTM model

Table 2. Accuracy Comparison among LSTM, Bi-LSTM and Bi-GRU model

Algorithm	Train Accuracy	Test Accuracy
LSTM	95.86%	76.29%
Bi -GRU	97.96%	76.10%
Bi- LSTM	97.96%	77.91%

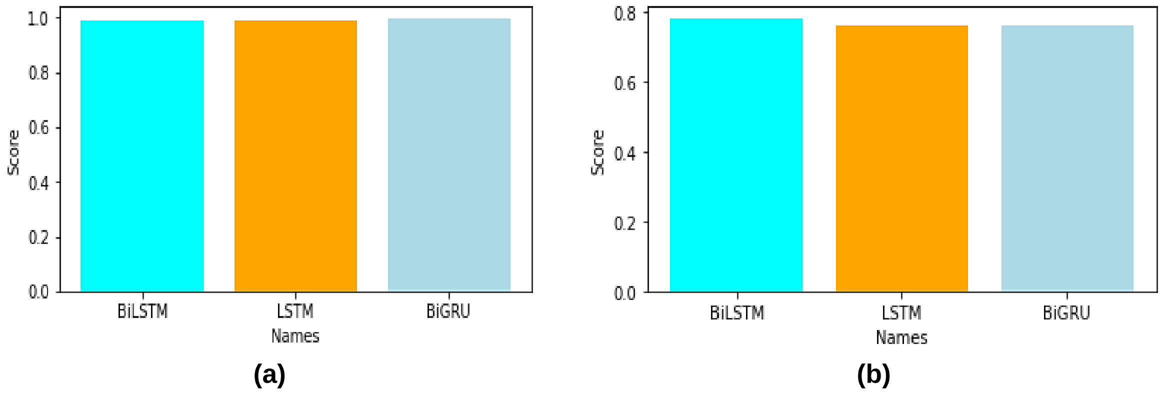


Fig. 24. (a) Training Accuracy (b) Testing accuracy

Table 3. Performance metrics comparison among LSTM, Bi-LSTM and Bi-GRU model

Algorithm	Precision	Recall	F1-Score
LSTM	76.23%	73.98%	73.52%
Bi- LSTM	76.15%	76.89%	76.66%
Bi -GRU	75.09%	75.91%	75.59%

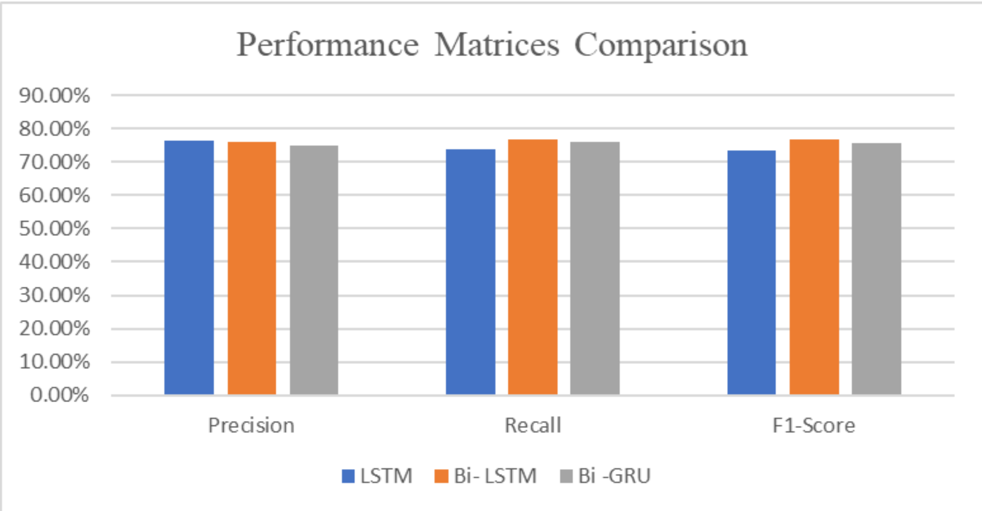


Fig. 25. Comparison chart of LSTM, Bi-LSTM, and Bi-GRU model for performance matrices

The comparison analysis results indicate that, out of the three deep learning models used for the newspaper headline classification research, the Bi-LSTM model performs the best. It has the best test precision (76.23%), sensitivity (76.15%), and accuracy (77.91%). This indicates that the Bi-LSTM model is the best at preventing false positives and false negatives while also being the most accurate model for predicting the right category of a newspaper headline.

The Bi-LSTM model is a member of the RNN family and was designed specifically to be an excellent model for capturing long-term dependencies in sequential data. Because of its special skill, it's quite good at text categorization jobs. For example, it can identify the category a headline belongs to by identifying patterns in the words and phrases in the headline.

Both LSTM and Bi-GRU, the other two deep learning models used in the research, are classified as recurrent neural networks (RNNs). When compared to the Bi-LSTM model, they do, however, show limitations in terms of their applicability for text categorization tasks. Specifically, the LSTM model is better at capturing short-range dependencies than the Bi-

GRU model, which is good at capturing long-range dependencies but not as well as the Bi-LSTM model.

To summarize, the research results indisputably demonstrate that the Bi-LSTM model is the best deep learning technique for classifying newspaper headlines. It is the most accurate in predicting the right headline category and also performs exceptionally well at reducing false positives and false negatives. This crucial result has important ramifications since it shows how the Bi-LSTM model might support the creation of more accurate and effective methods for classifying newspaper headlines. These kinds of solutions could make it easier and faster to get interesting news, which would help readers find pertinent information more quickly. News organizations can also use these developments to better understand the interests and preferences of their viewers, which will improve the way news is delivered in general.

5. CONCLUSION

It is challenging to create a classification system for Bengali news headlines that is especially efficient. Bengali language is rich and varied in a multitude of aspects. The Word2Vec method is examined. By design, the dataset includes most of the words that are relevant to the classes. In this study, you looked into the difficulty of identifying Bengali newspaper headlines using LSTM, Bi-LSTM, and Bi-GRU.

The models were evaluated and trained using a dataset comprising primary and secondary data of Bengali newspaper headlines, and the results were analyzed. The aforementioned analysis shows that Bi-LSTM performs better than the other two deep learning techniques. Its classification accuracy for Bengali newspaper headlines is 97.96% for training and 77.91% for validation; its corresponding f1 score, recall, and precision scores are 76.15%, 76.89%, and 76.66%. The Bi-GRU and LSTM models offer the second-best training accuracy, coming in at 97.28% and 95.86%, respectively. The accuracy of validation is almost 77%. The employed small dataset is the reason why the findings are not even close to 100%. We believe that our proposed approach will yield good results with a larger dataset.

As the volume of data increases, algorithms exhibit varying behaviors. We have also tested with multiple hyperparameters to find the optimal combination that yields the highest performance on the validation set. Despite our best efforts, we have found that the models continue to show overfitting, which suggests that further work is needed to improve their performance. Beyond the technical challenges we faced during this inquiry, we encountered issues with the nature of the dataset. We found that the dataset is rather limited and lacks diversity, which could limit the ability of the models to generalize to new data. We have attempted to tackle this issue by the application of data augmentation techniques; nonetheless, we believe that greater performance would necessitate a more extensive and diverse dataset. To solve this problem more effectively, we want to use a larger primary dataset in our future work.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Wang C, Nulty P, Lillis D. A comparative study on word embeddings in deep learning for text classification, in Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval, New York, NY, USA: ACM. Dec. 2020;37–46. DOI:10.1145/3443279.3443304.
2. Shahin Md, Ahmmed Tanvi, Piyal, Shahriar, Shopon, Md. Classification of Bangla News Articles Using Bidirectional Long Short-Term Memory. 2020;1547-1551. DOI:10.1109/TENSYMP50017.2020.9230737.
3. Cai J, Li J, Li W, Wang J. Deep learning Model Used in Text Classification. 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, China. 2018;123-126. DOI:10.1109/ICCWAMTIP.2018.8632592.
4. Hossain MR, Hoque MM, Sarker IH. Text classification using convolution neural networks with FastText Embedding, in Advances in Intelligent Systems and Computing, 2021. DOI:10.1007/978-3-030-73050-5_11.
5. Rana MI, Khalid S, Akbar MU. News classification based on their headlines: A review, in 17th IEEE International Multi

Topic Conference: Collaborative and Sustainable Development of Technologies, IEEE INMIC 2014 – Proceedings; 2015. DOI:10.1109/INMIC.2014.7097339.
6. Hossain A, Chaudhary N, Hasan Rifad Z, Hossain BMM. Bangla News Headline Categorization. Int. J. Educ. Manag. Eng. 2021;11(6):39–48. DOI:10.5815/ijeme.2021.06.05.
7. Khushbu SA, Masum AKM, Abujar S, Hossain SA. Neural network based bengali news headline multi classification system: selection of features describes comparative performance, 2020 11th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT. 2020;2–7. DOI:10.1109/ICCCNT49239.2020.9225611.
8. Kaur Gurmeet, Bajaj Karan. News Classification using Neural Networks. Communications on Applied Electronics. 2016;5:42-45. DOI:10.5120/cae2016652224.
9. Zhang Y, Rao Z. N-BiLSTM: BiLSTM with n-gram Features for Text Classification Proc. 2020 IEEE 5th Inf. Technol. Mechatronics Eng. Conf. ITOEC. 2020;1056–1059 DOI:10.1109/ITOEC49072.2020.9141692.
10. Akanksha Patro, Mahima Patel, Richa Shukla and Dr. Jagurti Save. Real Time News Classification Using Machine Learning. International Journal of Advanced Science and Technology. 2020; 29(9s):620-630. Retrieved from http://sersc.org/journals/index.php/IJAST/article/view/13254
11. Hoq Muntasir, Haque Promilax, Nazim Uddin, Mohammed. Sentiment Analysis of Bangla Language Using Deep Learning Approaches; 2021. DOI: 10.1007/978-3-030-76776-1_10.
12. Ahmed A, Yousuf MA. Sentiment Analysis on Bangla Text Using Long Short-Term Memory (LSTM) Recurrent Neural Network. In: Kaiser MS, Bandyopadhyay A, Mahmud M, Ray K. (eds) Proceedings of International Conference on Trends in Computational and Cognitive Engineering. Advances in Intelligent Systems and Computing. 2021;1309. Springer, Singapore. DOI:10.1007/978-981-33-4673-4_16
13. Roy, Amartya, Sarkar Kamal, Mandal Chintan. Bengali Text Classification: A New multi-class Dataset and Performance Evaluation of Machine Learning and Deep Learning Models; 2023. DOI: 10.21203/rs.3.rs-3129157/v1.
14. ChenJX, Jiang DM, Zhang YN. A Hierarchical Bidirectional GRU Model with Attention for EEG-Based Emotion Classification. IEEE. 2019;7:118530–118540. DOI:10.1109/ACCESS.2019.2936817.
15. Siam ASM, Hasan MM, Talukdar MM, Arafat MY, Jobayer SH, Farid DM. Bangla News Classification Employing Deep Learning. In: Thai-Nghe, N., Do, TN., Haddawy, P. (eds) Intelligent Systems and Data Science. ISDS 2023. Communications in Computer and Information Science, vol 1949. Springer, Singapore. 2023. DOI:10.1007/978-981-99-7649-2_12
16. Chowdhury Pallab, Eumi Ettilla, Sarkar Ovi, Ahamed Md. Bangla News Classification Using GloVe Vectorization, LSTM, and CNN; 2022. DOI:10.1007/978-981-16-6636-0_54.
17. Bharath G, Manikanta KJ, Prakash GB, Sumathi R, Chinnasamy P. Detecting Fake News Using Machine Learning Algorithms. Int. Conf. Comput. Commun. Informatics, ICCCI. 2021;1–5. DOI: 10.1109/ICCCI50826.2021.9402470.
18. Habib Md. Ahsan, Akter Asma. Deep learning Bangla text classification using recurrent neural network. 2022;8:10-16.
19. Prothom Alo; 2022. Available:http://www.prothomalo.com
20. The Daily Ittefaq; 2022. Available:https://www.ittefaq.com.bd
21. Bangladesh Pratidin; 2022. Available:https://www.bd-pratidin.com
22. Kaler Kantho; 2022. Available:https://www.kalerkantho.com
23. The Daily Inquilab; 2022. Available:https://dailyinqilab.com
24. Jugantor; 2022. Available:https://www.jugantor.com
25. Wang C. NultyP, Lillis D. A comparative study on word embeddings in deep learning for text classification, ACM Int. Conf. Proceeding Ser. 2020;37–46. DOI:10.1145/3443279.3443304.
26. Rahman, Shadikur, Hossain, Syeda, Islam, Saiful, Chowdhury, Mazharul, Rafiq, Fatama & Biplob, Khalid; 2019. Context-Based News Headlines Analysis Using Machine Learning Approach. DOI:10.1007/978-3-030-28374-2_15.

27. Available:https://www.sciencedirect.com/topics/computer-science/one-hot-encoding, (2023)
28. Oriol Vinyals, Understanding LSTM Networks, Colah. GitHub; 2015. Available:https://colah.github.io/posts/2015-08-Understanding-LSTMs,
29. Schuster M, Paliwal KK. Bidirectional recurrent neural networks, in IEEE Transactions on Signal Processing. Nov. 1997;45(11):2673-2681. DOI:10.1109/78.650093.
30. Available:https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be, (2023)
31. Ting KM. Confusion Matrix. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA; 2011. DOI:0.1007/978-0-387-30164-8_157.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/110372