A Study Towards Bangla Fake News Detection Using Machine Learning and Deep Learning

Chapter · October 2021

DOI: 10.1007/978-981-16-5157-1_7

Elias Hossain
North South University

Md Nadim Kaysar
Nagoya Institute of Technology

Abu Zahid Md Jalal Uddin Joy
Daffodil International University

Md. Mizanur Rahman
Rajshahi University of Engineering & Technology

A Study towards Bangla Fake News Detection Using Machine Learning and Deep Learning

Md. Elias Hossain[1], Md. Nadim Kaysar[2], Abu Zahid Md Jalal Uddin Joy[1], SM Mizanur Rahman[3], Md. Wahidur Rahman*

[1]Department of Software Engineering, Daffodil International University,

[2]Department of Computer Science & Engineering, World University of Bangladesh,

[3]Department of Computer Science and Engineering, Rajshahi University of Engineering and Technology

*Department of Computer Science & Engineering, Mawlana Bhashani Science and Technology University, Bangladesh

wahidtuhin0@gmail.com

Abstract. Verifying Bangla Fake news is challenging, especially if there are many updates from various sources such as social media or online news portals. This study aims to identify the Bangla fake news article; therefore, our Corpus is trained with 57000 Bangla news items related to trustworthiness and counterfeit. In this study, 95% and 94% accuracy was found by applying K-fold cross-validation on top of Bi-LSTM with Glove and FastText model. At the time, the research is also experimented with the state-of-the art technique like Gated Recurrent Unit (GRU) and found the accuracy 77%. In the sharp contrast, we tracked out the accuracy of 96% utilizing the Bi-LSTM which identically indicates our proposed model. A comparative study on existing work has been utilized in this research. Again, some experimental analysis based on the FEM is shown elaborately in this research. However, the proposed system can be adjustable in real-time news classification of Bangla Fake news.

Keywords: Bangla Fake News, Text Classification, Machine Learning, Random Forest, LSTM, Bi-LSTM, CNN, Glove, Fasttext, Gated Recurrent Unit (GRU)

1 Introduction

The world today is rich with data and information on news and events. Due to modern communication technology, people today have easy access to information. This also means data is also produced in large amounts by millions who have access to the internet, especially social media. Access to data is not a problem, but it is dangerous when these data include fake news or stories that spread worldwide. This fake news can cause of misunderstanding and even political tension and it has immoral impression in personal and social life. A recent report illustrates, about one in three citizens in the United States, Spain, Germany, United Kingdom, Argentina, and South Korea claim they have seen untrue or deceiving data on social media related to COVID-19 [1].

Impact of fake news is producing devastation worldwide. In 2012, several Pagoda was torched in Bangladesh after an image showed a derogatory picture of the Quran. The image was tagged with a Buddhist youth but was not responsible for the photo in any way. Many rumours spread because of this image even though the image's identity could not be verified [2]. In 2019, a false rumour circulated online about the Padma Bridge authorities sacrificing humans' lives at the construction site. This rumour then leads to suspicion of random individuals being child kidnappers [3]. In Bangladesh, the 2012 Ramu incident is a standard example where almost 25 thousand people joined in abolishing the Buddhist temples on the source of a Facebook post from a fake account [4].

Fortunately, there are website to tackle fake news, But these websites are not able to sufficiently respond to many fake news incidents. There some computational approaches used to isolate fake news, which is destructive for our daily life. Nowadays, we found lots of research work in English Languages. As of now, more than 341 million people in the world communicate with Bangla Language. But there is no resource to detect fake news written in the Bangla language. The aim of this work is to develop a solution to fight against misinformation. This study also aims to create an AI-based effective system to identify fraud news article by applying Traditional Machine Learning and Deep Learning algorithms. Through our proposed solution, we can easily classify inaccurate Bangla news. Our contribution can be summarized follows:

- We have presented a comparative analysis based on numerous feature extraction approach with traditional machine learning and deep learning.
- We develop a classification system to detect fake news written in the Bangla language. Additionally, we have implemented different type’s pre-trained model in this problem.
- We have shown a separate text preprocessing pipeline for each approach, which will positively impact the research community.

The manuscript is classified into five sections. Section two presents the backgroud study of previous contributions. Section three shows the proposed methodology and working principles. Section four illustrates the results with proper discussion. Finally, section five depicts the conclusion of this manuscript.

2 Related Works

The author of paper [5] proposed the multinomial Naïve Bayes model to detect Malicious Bangla Text Content by using social networks. The proposed research recognizes spam based on the extremity of each sentence related to it, and accuracy was found 82.44 %. [6] Was tried to detect hateful speech in Public Facebook Pages for the Bengali Language. The authors developed a machine learning-based model, but they could not achieve satisfactory accuracy. So a neural network-based Gated Recurrent Unit (GRU) model was considered. The accuracy of the GRU was 70%.

[7] Conducted research for the Automatic Detection of Satire on Bangla Documents. Convolutional Neural Network (CNN) was used for this model and obtained 96.4% accuracy. [8] Used a Deep Convolution Nets approach to categories the Bangla Documents programmatically. Word2vec was utilized for feature extraction, and they proposed a "DCNN" deep learning-related neural network, whose accuracy was 94%. [9] Proposed Bangla word embedding and its application of Bangla text classification. The Skip n-gram method was applied for word embedding with various dimensions and Support Vector Machine (SVM). [10] Showed Boosting on Stylometric and Word Vector Features method to identify fake news with 95% accuracy. [11] Demonstrated their work by using Convolutional Neural Network (CNN) for the detection of fake news. The study was conducted on the English dataset, and the performance score stood at 98%. [12] Classified fake news by using the LSTM approach. The authors used different dens layers and various percentages of filters using the Conv1D layer. Glove pre-trained word embedding to find out fraud news. [13] Researched detecting fake news on Social Media Networks. Fake news was classified by applying various types of traditional machine learning algorithms. [14] Explained tensor decomposition-based deep neural network for improving fake news detection. A deepFake method was proposed by applying the BuzzFeed and PolitiFact dataset. In [15] research proposed a CNN based flood management system, which is predict available space of the tank using IOT sensors data. [16] Conducted with traditional machine learning approach with imbalance data and they are used numerous types of techniques resolve imbalance data problem.

Reviewing the above study shows that not enough research has been accomplished on vast datasets. Some drawbacks can be found in the above study mentioned. Authors [5] acknowledges they do not have sufficient Corpus for their work. The authors [6] detect hateful speech from Public Facebook Pages for the Bengali Language over 5,126 Bengali comments; they have only experimented through traditional machine learning algorithms. Still, the amount of datasets is comparatively less.

In our proposed research, we have applied deep learning and machine learning algorithms for classifying fake news reports in the context of Bangla language. Various features have been generated for the conventional approach, such as unigram and bigram. This is the most extensive dataset in Bangla, with a volume of 57000 as far as we know and handling such a massive dataset is exceptionally challenging. We have shown a separate text preprocessing pipeline for each approach, which will positively impact the research community. While not enormous research has been done on top of Bangla, it will serve as a Benchmark for those who desire to work with Bangla News Classification.

3 Proposed Research Methodology (PRM)

In this proposed study, several traditional machine learning and deep learning algorithms are tested in terms of Bangla misleading news identification from news reports. The Proposed Research Methodology (PRM) is divided into four segments:

Experimental Setup, Data Preparation Pipeline (DPP), Features Extraction Methods (FEM), and Algorithm Selection Procedure (ASP).

By looking at Fig.1, it can be observed that the proposed research architecture diagram is classified into several phases, and each stage complete an individual task. More importantly, text preprocessing is completed in the first stage using pre-trained and non-pre-trained models; then, the model is fed into the machine learning and deep learning algorithms; afterwards, the model evaluation is accomplished, moving to architecture embodiment eventually, the documents are being classified. The details sequence and consequences are visualised in Fig.1.

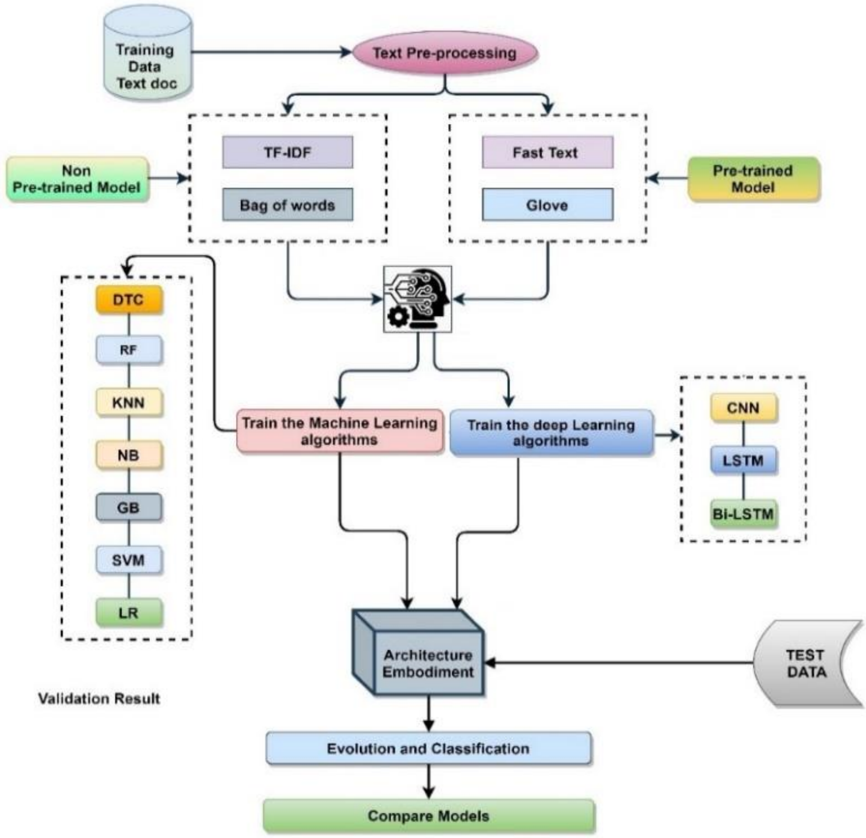

Fig. 1. Architecture diagram of the proposed research.

```
BEGIN   | Step 1: Data read from CSV file
        | Step 2: Data pre-processing
        | Step 3: Feature Extract:
        |           if(Machine Learning Based Feature Extract):
        |              Use TF-IDF|| Bag of Words || Word2Vec method
        |          else:
        |              Use One hot encoding || Glove/fastText pre-trained
```

```
        Step 4: Model Selection:
                  If(Machine Learning based algorithm):
                     Use DTC || RF ||SVM ||NB||GB||LR||KNN
                  Else:
                      Use LSTM || CNN|| Bi-LSTM|| GRU
END     Step 5: Output of the model.
```

Psedo code.1. Working flow of propose approach.

3.1 Experimental Setup

In this section, we have described our research dataset based on the Bangla news article. The proposed research dataset contains 57000 news articles that were collected from various news portals of Bangladesh. We have collected the dataset from [4]. Table No 1 shows the dataset description.

Table 1. News description.

Feature	Authentic	Fake
Domain	jagonews24.com	channeldhaka.news
Publication Time	2018-09-19 22:00:27	2019-02-22T14:50:20+00:00
Category	National	Technology
Headline	রোকসানাকে দেখলেই আঁতকে উঠবেন যে কেউ	হারানো মোবাইল খুঁজে পাবার সহজ উপায়
Article	১২ বছর বয়সী রোকসানার জীবন দুর্বিসহ করে তুলেছে রাজধানীর ওয়ারী এলাকার পাষণ্ড এক দম্পতি।	মোবাইল হারিয়ে যওয়াটা কোনো জটিল ব্যাপার নয়। হারানো মোবাইল খুঁজে পাবার নিয়মটাও বেশ সহজ। ম...
Label	1	0

3.2 Data Preparation Pipeline (DPP)

Data preparation is an essential part of applying every machine learning algorithm. The same with textual data before using machine learning algorithms rules to the text data requires data preparation [17]. Data Preparation concludes with data preprocessing. Text preprocessing recommends stops words, punctuation, phrases that do not carry much weight in context to the text, etc.

Noise Filtering &Text Normalization. In this section, we have filtered noises from our text corpus, such as removing constant features, dropping duplicate data, dropping identical columns, removing null values, applying regular expression, and removing stop words. Bangla is still now as a low resource has not yet developed a large corpus like NLTK, but some stop words are found in the case of Bengali text data, for example, অবশ্য, অনেক, অনেকে, অনেকেই, মধ্যভাগে, যাদের, যাচ্ছে,দিয়েছে, শুধু, সেটাও, মধ্যভাগে and so on. We have removed Stop Words through BLTK Library [18]. Stemming programs are known as stemming or stemmers. An algorithm reduces the words "chocolates", "chocolatey", "choco" to the foundation word, "chocolate" and "retrieval", "retrieved", "retrieves" cut back to the stem "retrieve". We have endeavored for Stemming through the BLTK library [18].

3.3 Features Extraction Methods (FEM)

In this section, Feature extraction approaches are described. Features must be extracted to apply machine learning algorithms since the machine only understands numerical data. This section has been classified into Non Pre-trained Model Intuition (NPMI) and Pre-trained Model Intuition (PMI).

Non-Pre-trained Model Intuition (NPMI).

Term Frequency (TF) — Inverse Document Frequency (IDF). In Natural Language Processing, various features extraction techniques can be found to extract features for the text document; for instance, Bag of Words (BOW), Term-Frequency-Inverse-Document (TF-IDF), Count Vectorizer, Word2vec, etc. [19]. TF (Term Frequency), which measures how frequently terms occurred in a document. On the other hand, Inverse Document Frequency (IDF) is a score of words in a full document. Another significant parameter of the TF-IDF feature extraction approach is N-gram that is a sequence of N tokens or words [20]. We can write the equation of TF-IDF as follows:

$$TF: tf(w,d) = \log\big(1 + f(w,d)\big), IDF: tf(w,D) = \log\left(\frac{N}{f(w,D)}\right) \quad (1)$$

$$TF\ IDF: tf - idf(w,d,D) = tf(w,d) * tf(w,D) \quad (2)$$

Over here, w is a word, and d is a document. *F(w,d)* is a frequency of the word in a document-the inverse document frequency (IDF) of the word across a set of documents. Here N is the number of documents and *f(w,d)* is a number of the document containing the corpus word. Multiplying these two numbers results in the TF-IDFscore of a word in the document.

Pre-trained Model Intuition (PMI)

Word2Vec. Word2Vec is a superficial word embedding model proposed by [21]. Word2Vec describes word embedding with two-layer shallow neural networks. It has several conveniences over the bag-of-words and the TF-IDF approach. It takes on the semantic meaning of the document's different words. It has two types of architecture inside CBOW (Continuous Bag-of-Words) and skip-gram model [22]. We proposed a skip-gram model and skip gram uses word information to predict its neighbor word, which defined as:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c, j\neq 0} log\ p(\ \omega_{t+j}|\omega_t \quad (3)$$

In the above equation, c denoted as a training context (c become a function of the center word ω_t).The besic formula of skip-gram by using the softmax function:

$$p(w_O \mid w_I) = \frac{\exp\left(v_{w_O}^{\top} v_{w_I}\right)}{\sum_{w=1}^{W} \exp\left(v_w'^{\top} v_{w_I}\right)} \quad (4)$$

Where ϑ_{ω} is input and ϑ'_{ω} is the output of the vector representation of ω. W is the number of words in the vocabulary. The cost computing $\nabla \log(\omega_o|\omega_i)$ is proportional to W, so that formula is impractical.

FastText & Global Vectors for Word Representation (Glove). FastText is an extension of word2vec model. It treats each word as the composition of character n-grams. So, the vector for a word is made of the sum of this character n grams. It generates better word embedding's for rare words even if words are rare their character n grams are still shared with other words. Glove treats each word in corpus like an atomic entity and generates a vector for each word. It aims to conciliate the word prediction models with the word statistics over a whole corpus. They provide a model that considers co-occurrence statistics of the corpus and the efficiency of prediction-based methods.

3.4 Algorithm Selection Procedure (ASP)

In our proposed research, we have experimented through the Supervised Machine Learning Algorithm and Deep Learning Algorithm. We have classified this section into two phases: the Machine Learning Model (MLM) and Neural Network Model (NNM).

Machine Learning Model (MLM). In this section, six classification algorithms applied, such as Decision Tree Classifier, Random Forest Classifier (RF), K Nearest Neighbor (KNN), Multinomial Naïve Bayes, Gradient Boosting, Support Vector Machine (SVM), and Logistic Regression. After experimenting, we observed that the Decision Tree Classifier (DTC) and Random Forest Classifier (RF) performed well, so the DTC is described in this section.

Decision Tree Classifier (DTC). Decision Tree Classifier (DTC) is a prior classification algorithm for text and data mining [23]. DTC is used effectively for classification in various fields [24].The primary concept is making a tree that supported the attribute for categorized data points. However, the biggest challenge of a DTC is that attribute or feature might be at the parents' level and that one ought to be at the kid level. An applied math modelling [25] addressed feature choice within the tree to solve this downside. For a training set comprising p positive and n negative:

$$H\left(\frac{p}{n+p}, \frac{n}{n+p}\right) = -\frac{p}{n+p}\log_2\frac{p}{n+p} - \frac{n}{n+p}\log_2\frac{n}{n+p} \quad (5)$$

Choose K in the attribute with the unique value, the training set E divides into the prefixes of $\{E_1, E_2, \ldots, E_k\}$ The expected entropy (EH) will remain after the attempt in the attribute (branches $i = 1,2, \ldots, k$) including:

$$EH(A) = \sum_{i=1}^{K} \frac{p_i + n_i}{p + n} H\left(\frac{p_i}{n_i + p_i}, \frac{n_i}{n_i + p_i}\right) \quad (6)$$

Information gain (I) or decrease in entropy for this trait is:

$$A(I) = H\left(\frac{p}{n+p}, \frac{n}{n+p}\right) - EH(A) \quad (7)$$

Select the property with the most prominent information gain as a parent's hub.

Neural Network Model (NNM). This section explains Convolutional Neural Networks (CNN), Long-short-term memory (LSTM), and Bidirectional LSTM (BI-LSTM) that classify the text. CNN consists of five structures: Convolution layer, pooling layer, fully connected layer, Dropout, and Activation function.

Convolutional Neural Network (CNN). The basic functionality of a convolution neural network is similar to that of the animal brain's visual cortex. Convolution neural networks perform well in text classification tasks. The criteria for text classification are identical to those for image classification, except that instead of pixel values, we have a matrix of word vectors. Convolutional Neural Networks (CNN) is one of the most popular algorithms for machine learning. The CNN illustrates fruitfulness to classify the short and long text [26]. So we examined categorizing Bangla as Fake news with the help of CNN. CNN's hidden layers usually contain convolutional layers, pooling layers, fully connected layers, and generalization layers [27]. Here, embedding represents every word numeric vector, and the embedding creates a data table where each row of that table representing the embedding of a word. Every embedding have some parameters like voc-size and embedding-dim. Here voc_size represent the number of unique words in a document and embedding-dim, describing the number of dimensions of each term [11].

Convolutional Layer. This is the first layer and one of the primary components of a Convolutional Neural Network (CNNs). They take as input the raw pixel values of the training image and extract features from them. This layer ensures that pixels are spatially related by learning image features from small squares of input data [28].

Mathematically, we can define as a combination of two functions "f" and "g" as:

$$(f * g)(i) = \sum_{j=1}^{m} g(j).f(i - j + \frac{m}{2}) \quad (8)$$

Pooling Layer. Another essential concept of convoluted neural networks is pooling, recognized as a form of non-linear down-sampling. Pooling performs to extract the Particular Value (Max / Average Value) from the specified portion (n × n) of the matrix. In this architecture, the max-pooling layers combine the output from the convolutional layers. All results from the convolutional layers are concentrated and conveyed to the following level of the convolutional layer [29].

Fully Connected Layer. The fully connected layer is also known as the dense layer. Each neuron receives input from its predecessor neuron, and it is densely connected. The output calculates by optimizing the loss with the result obtained by weight initiation with each neuron [30]. Suppose we have an input layer where the inputs are x1, x2, and x3 and a hidden layer H1. The weight initialization of each neuron is y1, y2, and y3. After adding bias, we can see the output following:

$$Y = \sum_{i=1}^{n} wi * xi + wn * xn \ldots \ldots \ldots \ldots \ldots + bias \qquad (9)$$

Two types of propagation can be used to reduce errors, such as forward propagation and backward propagation. When considered multiple hidden layers, in that case, the error usually reduces through the backpropagation. In backpropagation, derivatives are complete by chain rules [31].

Dropout. The Dropout, which helps avoid overfitting, randomly sets input units to 0 with a frequency of rate at each stage during training time. It should be noted that the Dropout is only active when training is set to True, which means that no values are lowered during inference when using the model.

Activation Function. We have considered the ReLu as an activation function for our CNN architecture. The foremost advantage of the ReLu function is that it eliminates the negative rate from an activation map by labeling them zero in a network. The vanishing gradient problem is solved efficiently through this function [31].

Long Short Term Memory (LSTM). LSTM's is a solution to short-term memory. They have inner mechanisms described gates that can regulate the flow of information. These gates can learn which data in a sequence is important to keep or throw away. By doing that, it can pass relevant information down the long chain of flow to make predictions. Almost the state of the art result based on recurrent neural network is achieved with these networks. Having the capability to catch up with the sequential information, the LSTM usually widely used for text classification-related issues. In particular, by collecting sequential information from both directions in texts, Bidirectional LSTM (Bi-LSTM) has demonstrated impressive efficiency. Besides, when used with Bi-LSTM, the attention mechanism has been noticed as a potential pooling technique for classification tasks. In this research, we have experimented through Bidirectional LSTM (Bi-LSTM) model that is close to focus on top [32]. However, the following architecture that has been used in Bi-LSTM: Input contains vocabulary size, embedding_vector_features and input shape, SpatialDropout1D layer with value 0.4, Bidirectional with 356 LSTM units, Dropout with value 0.2 to avoid overfitting, A dense layer with 2 neurons with softmax activation function.

4 Result and Analysis

This section consists of the two-phase, to illustrate, Performance Analysis (PA), Comparative Analysis (CA).The PA further subdivided into two-part, for instance, Consequence of the Model (CM) and Model Evaluation Report (MER).

4.1 Performance Analysis (PA)

This section mainly analyses the performance obtained through the classification algorithms of machine learning. However, two approaches are described in this section: the performance of the traditional machine learning algorithm and the deep learning algorithm. Our research found that Deep Learning algorithms play a vital role in Text classification rather than Traditional Machine Learning Algorithms.

Consequence of the Model (CM). As we have mentioned earlier, this study has experimented through the seven classification algorithms, so the precision, recall, and f1-score of the algorithms are shown in Table No 2. In Tables 3 and 4 where the precision has been written "P" and the same way the recall is "R" and the F1-Score is "F1". We are present number of TP, TN, FP, and FN by confusion matrix. Table No 2, 3 and 4 shows the Classification report of the traditional approach and Deep Learning-based approach. Equations (10), (11), (12), and (13) show the formula for finding precision, recall, F-1 score, and accuracy. We obtained higher accuracy in the random forest (ensemble learning), which is 89% accuracy. Furthermore, On the other hand DTC,KNN,NB,GB,SVM and LR are gives 86%,78%,67%,73%,88% and 73% accuracy with Uni-gram. Table No 1 describes the same traditional algorithm with Bi-gram approach and got maximum 78% validity from Random Forest (RF). Besides, others algorithm are performance not so well. The word2vec method used in Table No 3 and got 83% accuracy in RF. Our Neural network model has shown outstanding performance, and two pre-trained models are also used to optimize the accuracy in Table No 4. In the LSTM experiment, we got accuracy of 96%. On the other hand, we conduct Bi-LSTM with validity of 96%. Here, CNN and GRU models are implemented with 95% and 78% accuracy, and the CNN model of authentic news F1 score is 96%, and Fake F1 score is 95%. Moreover, we have conducted the glove and Fasttext pre-trained model with Bi-LSTM, and the accuracy was 95% and 94%.

$$\text{precision} = \frac{TP}{TP + FP} \quad (10) \qquad \text{recall} = \frac{TP}{TP + FN} \quad (11)$$

$$F1 = 2 \cdot \frac{\text{precision.recall}}{\text{precision} + \text{recall}} \quad (12) \quad \text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

Table 2. Classification report for TF-IDF (Uni-gram and Bi-gram) approach.

Algorithm	Approach	Authentic			Fake		
		P	R	F1	P	R	F1
DTC	TF-IDF (Uni-gram)	0.88	0.83	0.86	0.84	0.89	0.86
RF		0.91	0.87	0.89	0.87	0.91	0.89
KNN		0.93	0.51	0.66	0.66	0.96	0.78
NB		0.67	0.67	0.67	0.67	0.67	0.67
GB		0.75	0.66	0.70	0.69	0.78	0.73
SVM		0.96	0.77	0.86	0.81	0.97	0.88
LR		0.73	0.71	0.72	0.72	0.74	0.73
DTC	TF-IDF (Bi-gram)	0.88	0.59	0.71	0.69	0.92	0.79
RF		0.90	0.61	0.73	0.70	0.93	0.80
KNN		0.90	0.47	0.62	0.64	0.94	0.76
NB		0.84	0.53	0.65	0.66	0.90	0.76

GB	0.84	0.21	0.33	0.55	0.96	0.70
SVM	0.93	0.54	0.68	0.67	0.96	0.79
LR	0.84	0.58	0.68	0.68	0.89	0.77

Table 3. Classification report for Word2Vec approach.

Algorithm	Authentic			Fake		
	P	R	F1	P	R	F1
DTC	0.81	0.80	0.81	0.80	0.82	0.81
RF	0.84	0.80	0.83	0.82	0.84	0.83
KNN	0.84	0.75	0.79	0.77	0.86	0.81
NB	0.56	0.52	0.54	0.54	0.59	0.56
GB	0.61	0.63	0.62	0.61	0.59	0.60
SVM	0.57	0.34	0.43	0.53	0.74	0.62
LR	0.56	0.53	0.54	0.55	0.57	0.56

Table 4. Classificati Classification report of Deep Learning algorithms (based on the features extraction approach)

Algorithm’s	Feature Extraction	Authentic			Fake		
		P	R	F1	P	R	F1
LSTM	One hot encoding	0.93	0.99	0.96	0.99	0.93	0.96
Bi-LSTM	One hot encoding	0.93	0.99	0.96	0.99	0.93	0.96
CNN	One hot encoding	0.93	0.99	0.96	0.98	0.93	0.95
GRU	One hot encoding	0.76	0.82	0.79	0.80	0.74	0.77
Bi-LSTM with Glove	Glove	0.93	0.98	0.95	0.97	0.93	0.95
Bi-LSTM with Fasttext	FastText	0.93	0.96	0.94	0.96	0.92	0.94

Model Evaluation Report (MER). Another approach is to decide how good the performance of different classification models is the ROC-AUC curve. Fig 2 and 3 show the ROC-AUC curve on top of Bi-LSTM with Glove and Fasttext pre-trained model. We have used K=5 to cross validation. Here, K=1 fold is test data and K=2, K=3, K=4 and K=5 fold data are training data. ROC-AUC mean value of every fold Glove pre-trained model is 99%.On the other hand, ROC-AUC mean value of Fasttext pre-trained model 98%. This study has been applied to the ROC-AUC curve on top of the Bi-LSTM with Glove FastText model. The blue line in Fig.2 and 3 is ROC, and the space below this ROC is AUC. The higher the value of ROC close to the value 1.0 in blue marked region refers the significance of the trained model. Nevertheless, Cross validation details sequence and a clear visualization are shown in Fig.2 and 3, respectively. On the other hand, the confusion matrix, accuracy assessment, and loss measurement approaches are shown in Fig 4, 5, 6, and 7, respectively

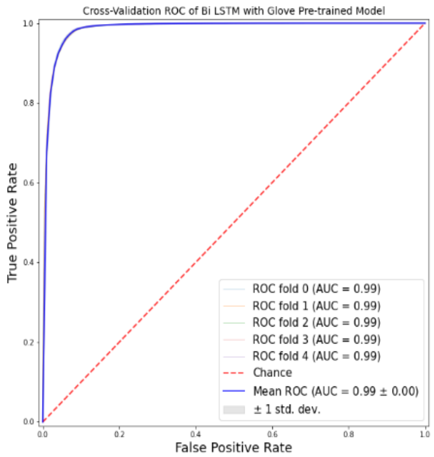

Fig. 2. Roc curve for Bi-LSTM.

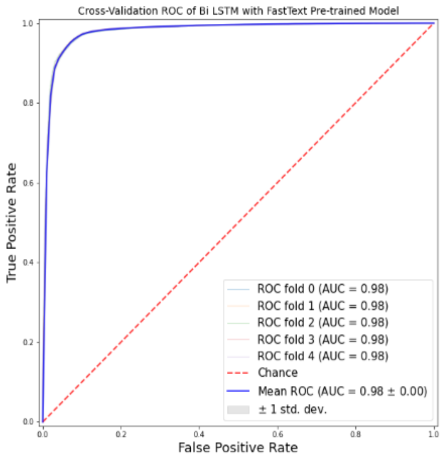

Fig.3. Roc curve for Bi-LSTM.

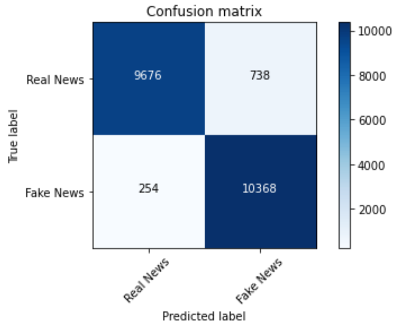

Fig. 4. Confusion matrix of Bi-LSTM with Glove.

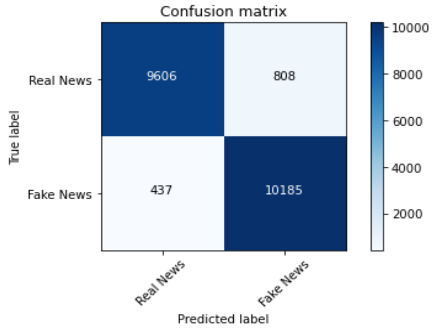

Fig. 5. Confusion matrix of Bi-LSTM with fastText.

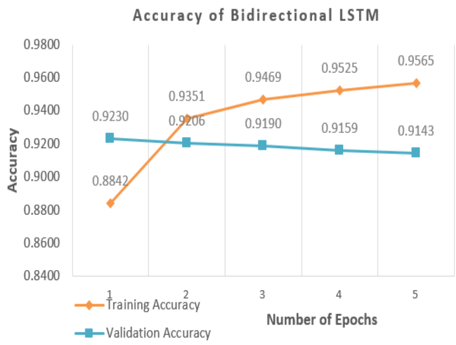

Fig. 6. Accuracy Mesurment of Bi-LSTM.

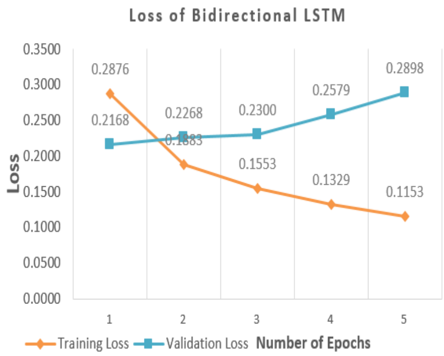

Fig. 7. Loss Mesurment of Bi-LSTM.

4.2 Comparative Analysis (CA)

This section discusses Comparative Analysis (CA). This research has been compared with the top four research papers that accomplished on Bangla News Classification. Table No.5 shows a comparison between recent publications with the proposed study. Articles in Table No 5 with higher accuracy than the proposed research are called "yes"; thus, papers with a lower accuracy refer to "no." If the previous study's accuracy equivalent with the proposed research, they are mentioned "equal."

Table 5. Comparasion Table.

Paper’s	Algorithm	Feature Extraction	Accuracy	Match with proposed accuracy (Yes/No/Equal)
[4]	MNB	Count Vectorizer, TF-IDF	82.44%	No
[5]	GRU	TF-IDF	70.10%	No
[6]	CNN	Word2Vec, TF-IDF.	96%	Equal
[8]	SVM	Word2Ve c	91%	No
Our proposed model.	Bi-LSTM	One Hot Encoding	96%	

5 Conclusions & Future work

Fake news contents are not only appears in English language but also seem in the others native languages. This research proposed a Machine Learning and Deep Learning approach with different features extraction pipelines that will place significant benchmark in the field of Bangla fake news classification. The previous contributors had experienced a higher accuracy on limited dataset, but this work had been utilized a dataset containing 57 thousand online articles. This research has tracked out accuracy of 96% by using the Bi-LSTM model. By experimenting with machine learning and deep learning algorithms, the proposed model has experienced the better performance with the deep learning models compared to the traditional machine learning algorithms. This paper also ensures a comparative study among existing works on Bangla fake news detection strategy. Though the proposed work has found the improved performance, it has experienced some limitations. First of all, Bangla language is still a very low resource language compared to the others existing languages. Secondly, there is no sufficient library as like NLTK to work with Bangla language. Finally, more and more data preprocessing are required to boost up the accuracy level of the proposed model. In the future, we will come up with a solution to crack these limitations to find the enhanced results in Bangla fake news detection and classification. However, the objective of this research is achieved and the proposed model can be practicable in real life Bangla fake news recognition.

References

1. Nielsen, R.K., et al., *Navigating the 'infodemic': How people in six countries access and rate news and information about coronavirus*. 2020: Reuters Institute.
2. Star, T.D. *A hazy picture appears*. 2012.
3. Star, T.D. *Mobs beat five dead for 'kidnapping'*. 2019.
4. Hossain, M.Z., et al., *BanFakeNews: A dataset for detecting fake news in bangla.* arXiv preprint arXiv:2004.08789, 2020.
5. Islam, T., S. Latif, and N. Ahmed. *Using social networks to detect malicious bangla text content*. in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*. 2019. IEEE.
6. Ishmam, A.M. and S. Sharmin. *Hateful speech detection in public Facebook pages for the Bengali language*. in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*. 2019. IEEE.
7. Sharma, A.S., M.A. Mridul, and M.S. Islam. *Automatic detection of satire in bangla documents: A cnn approach based on hybrid feature extraction model*. in *2019 International Conference on Bangla Speech and Language Processing (ICBSLP)*. 2019. IEEE.
8. Hossain, M.R. and M.M. Hoque, *Automatic bengali document categorization based on deep convolution nets*, in *Emerging Research in Computing, Information, Communication and Applications*. 2019, Springer. p. 513-525.
9. Ahmad, A. and M.R. Amin. *Bengali word embeddings and it's application in solving document classification problem*. in *2016 19th International Conference on Computer and Information Technology (ICCIT)*. 2016. IEEE.
10. Reddy, H., et al., *Text-mining-based fake news detection using ensemble methods.* International Journal of Automation and Computing, 2020: p. 1-12.
11. Kaliyar, R.K., et al., *FNDNet–a deep convolutional neural network for fake news detection.* Cognitive Systems Research, 2020. **61**: p. 32-44.
12. Agarwal, A., et al., *Fake news detection using a blend of neural networks: an application of deep learning.* SN Computer Science, 2020. **1**(3): p. 1-9.
13. Aldwairi, M. and A. Alwahedi, *Detecting fake news in social media networks.* Procedia Computer Science, 2018. **141**: p. 215-222.
14. Kaliyar, R.K., A. Goswami, and P. Narang, *DeepFakE: improving fake news detection using tensor decomposition-based deep neural network.* The Journal of Supercomputing, 2021. **77**(2): p. 1015-1037.
15. Smys, S., A. Basar, and H. Wang, *CNN based Flood Management System with IoT Sensors and Cloud Data.* Journal of Artificial Intelligence, 2020. **2**(04): p. 194-200.
16. Chakrabarty, N. and S. Biswas, *Navo Minority Over-sampling Technique (NMOTe): A Consistent Performance Booster on Imbalanced Datasets.* Journal of Electronics, 2020. **2**(02): p. 96-136.
17. Kalra, V. and R. Aggarwal. *Importance of Text Data Preprocessing & Implementation in RapidMiner*. in *ICITKM*. 2017.
18. Hossain, S., *BLTK: The Bengali Natural Language Processing Toolkit*. 2012.
19. Waykole, R.N. and A.D. Thakare, *A Review of feature extraction methods for text classification.* Int. J. Adv. Eng. Res. Dev, 2018. **5**(04).

20. Shah, F.P. and V. Patel. *A review on feature selection and feature extraction for text classification*. in *2016 international conference on wireless communications, signal processing and networking (WiSPNET)*. 2016. IEEE.
21. Mikolov, T., et al., *Efficient estimation of word representations in vector space.* arXiv preprint arXiv:1301.3781, 2013.
22. Naili, M., A.H. Chaibi, and H.H.B. Ghezala, *Comparative study of word embedding methods in topic segmentation.* Procedia computer science, 2017. **112**: p. 340-349.
23. Morgan, J.N. and J.A. Sonquist, *Problems in the analysis of survey data, and a proposal.* Journal of the American statistical association, 1963. **58**(302): p. 415-434.
24. Anyanwu Matthew, N. and G. Shiva Sajjan, *Comparative Analysis of Serial Decision Tree Classification Algorithms*. Issue.
25. De Mántaras, R.L., *A distance-based attribute selection measure for decision tree induction.* Machine learning, 1991. **6**(1): p. 81-92.
26. Shrestha, P., et al. *Convolutional neural networks for authorship attribution of short texts*. in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. 2017.
27. PAI, A., https://www.analyticsvidhya.com/blog/2020/01/first-text-classification-in-pytorch/. 2020.
28. Zhong, B., et al., *Convolutional neural network: Deep learning-based classification of building quality problems.* Advanced Engineering Informatics, 2019. **40**: p. 46-57.
29. Shang, L., et al. *Sentiment analysis of film reviews based on CNN-BLSTM-Attention*. in *Journal of Physics: Conference Series*. 2020. IOP Publishing.
30. Rakhlin, A., *Convolutional Neural Networks for Sentence Classification.* GitHub, 2016.
31. Bengio, Y., A. Courville, and P. Vincent, *Representation learning: A review and new perspectives.* IEEE transactions on pattern analysis and machine intelligence, 2013. **35**(8): p. 1798-1828.
32. Zhou, P., et al. *Attention-based bidirectional long short-term memory networks for relation classification*. in *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers)*. 2016.