2022 25th International Conference on Computer and Information Technology (ICCIT)
17-19 December, Cox's Bazar, Bangladesh

Bangla Fake News Detection using Machine Learning, Deep Learning and Transformer Models

Risul Islam Rasel [1,2], Anower Hossen Zihad [2], Nasrin Sultana[1,3], Mohammed Moshiul Hoque[1]
[1] Department of CSE, Chittagong University of Engineering & Technology, Chittagong-4349, Bangladesh
[2] Department of CSE, Chittagong Independent University, Chittagong-4000, Bangladesh
[3] Department of CSE, University of Chittagong, Chittagong-4331, Bangladesh
risul@ciu.edu.bd, zihadyeet@gmail.com, nasrin_cse@cu.edu.bd, moshiul_240@cuet.ac.bd

***Abstract*— News Categorization is one of the primary applications of Text Classification, especially, Fake news classification. In recent days, many researchers have done plenty of work on Fake news detection in rich resource languages like English. But, due to a lack of resources and language processing tools, research on low-resource languages like Bangla is still insignificant. In this study, we try to build a Bangla Fake news dataset combining newly collected fake news data and available secondary datasets. Previously available datasets contained redundant data, which we reduced in our experiment. Finally, we build a Fake news dataset that contains 4678 distinct news data. We experimented with our data with multiple Machine Learning (LR, SVM, KNN, MNB, Adaboost, and DT), Deep Neural Networks (LSTM, BiLSTM, CNN, LSTM-CNN, BiLSTM-CNN), and Transformer (Bangla-BERT, m-BERT) models to attain some state of the art results. The best performing models are CNN, CNN-LSTM, and BiLSTM, with the accuracy of 95.9%, 95.5%, and 95.3%, respectively. We also tested our models by applying the previously existing datasets, and we got a 1.4% to 3.4% improvement in accuracy from previous results. Besides accuracy improvement, our models show a significant increase in recall of fake news data compared to the prior studies.**

Keywords*— Bangla Fake News, Bangla Fake News classification, Bangla Text Classification, Bangla Natural Language Processing, Machine Learning, Deep Learning, Transformer.

I. INTRODUCTION

Fake news classification means separating real and fake news from the given news input text. Because of the internet, e-papers, online news portals, and social media have become substitutes for newspapers in many ways. The flow of information has never been this easy. This free flow of information comes at the cost of misinformation published by unauthorized sources. The intention is to establish ill-intended agendas or distasteful satire to create confusion, conflict, and unrest among mass people. In the worst cases, it can lead to violence. We have witnessed communal violence fueled by fake news on social media several times before, from Ramu [1] to Bhola [2]. There are many other instances [3] of violence and unrest sparked by fake news on social media or online news portals. Therefore, monitoring virtual space and filtering out fake news is a necessity.

Nowadays, text classification is a hot topic in Natural Language Processing (NLP). Over the past few years, researchers have been trying to extract significant features from unstructured text data for text classification purposes. Bag of Words, TF-IDF, and Word embedding (Word2Vec) are such kind of feature extraction techniques. The features are applied to train text classification models. Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) are two examples of deep learning algorithms which extract optimized features from unstructured textual data to categorize documents. Even though less in number, there have been some good works in Bangla language processing. However, most of the applications of the modern Deep Learning and Transformer models are still not being taken full advantage of mainly because of the scarcity of collected data and limitation of language processing tools. In the fake news classification category, there have been a handful of works using ML models but no mentionable work with DL and Transformers technology. Moreover, some previous work used imbalanced datasets and a limited amount of Fake news.

In this study, we experiment with traditional machine learning and deep learning models with and without an ensemble. We also use transformer models such as m-BERT and Bangla-BERT to check the dataset's generality. Finally, we perform a comparative analysis of the outcomes of all these models to determine the best-performing model with our experiment dataset. Additionally, we also perform a comparison with two previous benchmark datasets to validate our model and dataset generality. And we find that our proposed model and dataset perform better than the previous benchmark dataset. The contribution of our work is preparing a benchmark dataset by collecting new fake news data and combining previous Bangla Fake news datasets available online. We remove all the duplicate information from our dataset and finally prepare a new Bangla fake news corpus that contains 4678 unique news data with a 50:50 ratio of real and fake news. We applied state-of-art machine learning, deep learning, and transformer models to check our dataset's benchmarking. After a rigorous experiment, we see all ML, DL, and Transformer models able to classify fake news with more than 90% average accuracy, especially the highest recall (real Fake news) value is 95.9% in the LSTM model.

II. LITERATURE REVIEW

There are several types of fake news, such as satire, fabricated content, manipulated content, misleading content, and false content. In [4], the authors have categorized fake news into three categories, serious fake news, large-scale hoaxes, and humorous Fake. Muhammad Umer et al. [5] used a dataset containing four labels agree, disagree, discuss, and unrelated. These labels describe whether a news article is consistent with its headline. This type of analysis is handy for clickbait news categorization. Hussain et al. [6] used

979-8-3503-4602-2/22/$31.00 ©2022 IEEE

SVM and MNB classifiers to classify fake news from 2500 news data. They achieved up to 96.64% accuracy with SVM, whereas MNB got 93.32% accuracy. Tanvirul Islam et al. [7] experimented on 383 malicious Bangla comments collected from Facebook and Youtube. The MNB classifier used in their experimentation achieves 82.44% accuracy. Ishmam and Sadia [8] experimented with classifying six categories as Hate Speech, Communal Attack, Inciteful, Religious Hatred, Political Comments, and Religious Comments. They collected 5126 hateful speeches from Facebook. They achieved 52.2% accuracy from the Random Forest model and 70.10% accuracy using the GRU-based model. Shafayat et al. [9] applied Gaussian Naïve Bayes on a dataset consisting of 538 news headlines, of which half were Fake news headlines. This experiment attained 87% accuracy in Fake news headline classification. Farzana et al. [10] followed Muhammad Umer et al. [5] to categorize the type of fake news. They experimented on 726 articles, but they did not mention the number of Fake news in their dataset. They got up to 83% accuracy just training the headlines, whereas training on both headline and content gains 85% accuracy by the best performing model Random Forrest. Another study by Tasnuba et al. [11] used Passive Aggressive Classifier and SVM on a dataset prepared from two secondary datasets available online. Hossain et al. [12] experimented with multiple Machine Learning models with performance scores up to 96%. But that study suffered from a poor ratio of real and fake data. They used the dataset BanFakeNews [20] contained only 1176 fake news, whereas the number of real news was 38 times more. However, this study also explores DNN models such as LSTM, Bi-LSTM, CNN, and GRU. All these models have almost a similar precision score of 93%, whereas GRU scores 76%. Md. Zahin et al. [13] studied fake news detection based on multichannel combined CNN-LSTM on a dataset containing about 50k news. Their combined architecture of CNN-LSTM obtained up to 75.05% accuracy. Arnab et al. [14] worked to detect satire in Bangla documents with a CNN-based approach. They used a dataset containing 1480 satire news and gained 96.4% accuracy. M. M. Hoque et al. [15] conducted a study to categorize Bangla documents from 969,000 text documents with multiple Machine Learning, Neural Network, and Transformer models. They observed that their VDCNN + GloVe outperforms other models with an accuracy of 96.96%. Tanvirul et al. [16] experimented on various available corpus with Transformer models like BERT, XLM-RoBERTa base, and XLM-RoBERTa large. On the news dataset, they obtain 93.41% accuracy from the BERT model. Tanjim et al. [17] used BERT and ELECTRA on 44001 comments from Facebook posts to detect abusive Bangla comments. These transformer models gained up to 85% and 84.92% accuracy, respectively. M. Kowsher et al. [18] proposed the Bangla-BERT methodology, which has a 94.21% f1-score on the Banfakenews [20] dataset.

III. DATA

The authentic news data are abundant, but the scarcity of publicly available Bangla fake news collection makes the work challenging. Bangladesh's government has blocked almost all possible online fake news sources. We use the Internet Archive (https://archive.org/) to go back in the timeline and collect the Fake news data. Then we combine our collected data with the two publicly available datasets of Bangla news, BanFakeNews [20] and Bangla Fake-Real News [6] datasets. We have collected 500 Bangla fake news data via Internet Archive from ChannelDhaka, Earki, and Motikontho. These websites contain fake news of two sorts misinformation and satire. BanFakeNews [20] dataset contains 48678 real news from popular Bangla newspapers like Kalerkantho, Prothom Alo, etc and 1299 fake news from different news portals and e-papers such as Motikontho, Banglabeats, Bangaliviralnews, Shadhinbangla24, Prothombhor and more. However, among the 1299 fake news, there are 123 duplicate entries of the same news. Bangla Fake-Real News [6] dataset contains 993 fake news and 1548 real news from similar sources as the other datasets. This dataset contained 45 duplicate Fake news. Table I shows the data summary of our experiment data.

TABLE I. DATASET SUMMARY

Number of attributes	Category	Training data size	Test data size	Total	Ratio of Training and Test data
3	Real News	1754	585	4678	75 : 25
	Fake News	1754	585		

IV. METHODOLOGY

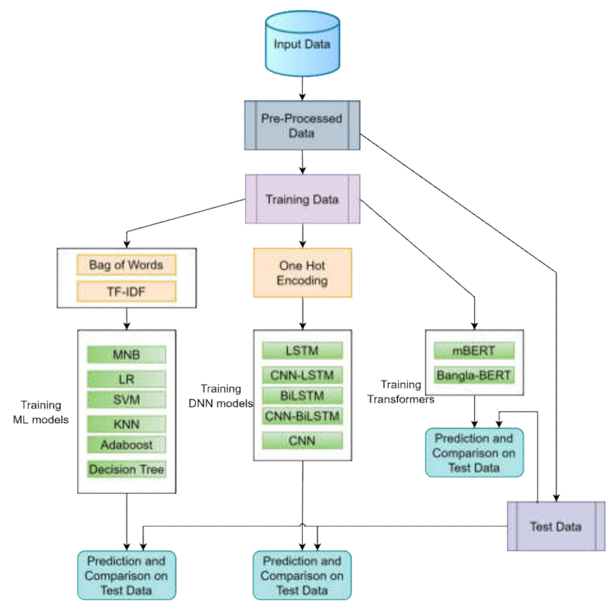

Fig. 1. Architecture of the proposed models

Fig. 1 shows the overall architecture of our experiment design. We applied state-of-art machine learning, deep learning, and transformer models to our proposed dataset and compared their results to check the dataset performance on the different model set up. Below we describe all the steps separately.

A. Data pre-processing

Text preprocessing is necessary to get better results from the classifiers. There may be unnecessary symbols, emoticons, or stopwords in text, which do not help the models in classification but increases the noise data. So, we remove unrelated columns, punctuations, special characters, and stopwords from the dataset. We applied the BNLP toolkit for text preprocessing purposes.

B. *Feature Extraction*

The feature extraction process helps to select optimized features to feed to the machine learning classifier algorithms as inputs. We used Count Vectorizer and TF-IDF Vectorizer (Term Frequency-Inverse Document Frequency) to extract numeric features from unstructured text data. Count Vectorizer converts a text into a vector based on the number of times each word appears across the full text. TF-IDF quantifies the significance of a word in a corpus or collection of documents. From unigram (single token), we get almost 86795 unique features.

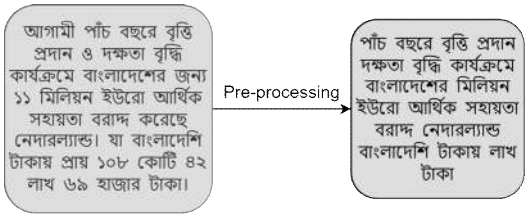

Fig. 2. Pre-processing sample

C. *Machine Learning Classifer Algorithms*

Six state-of-art machine learning classifiers such as Multinomial Naive Bayes (MNB), Logistic Regression (LR), K Nearest Neighbour (KNN), Support Vector Machine (SVM), Decision Tree (DT), and AdaBoost are applied to find out the performance of the news dataset. The parameters are examined and tuned to get optimized results. We set alpha value 1.0 for MNB model, used Limited memory BFGS (L-BFGS) with default penalty and c values for LR model, set neighbor (n) size to 12 for KNN, used RBF kernel and random state 7 for SVM model, used GINI impurity for the decision tree, and for Adaboost we set n estimators to 100 for getting best results.

D. *Deep Neural Netowork Classifers*

In this experiment, we used different Deep Neural Network (DNN) architectures such as Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory(BiLSTM), and a combination of CNN with LSTM (CNN+LSTM) and CNN with BiLSTM (CNN+BiLSTM) on our data. We used the One-Hot encoding technique to transform text data into numerical values. During the encoding, we used a maximum news length of 300 words, a vocabulary size of 86000, and pre-padding. We used adam optimizer with a learning rate of 0.001 for all models. We trained our models with 12 epochs with a batch size of 64. Early stopping (ES) is incorporated into the models to avoid overfitting. ES looks to improve validation accuracy with a min-delta of 0 and patience 3.

1) CNN: We used a three-layer Convolutional Neural Network (CNN) with three different kernel sizes, such as 4, 6, and 8. These layers use 32 filters each. We set a dropout layer with a 0.5 rate to skip overfitting and a max-pooling layer with the pool size 2 for downsampling. Lastly, the ReLU activation function adds non-linearity, and the probability distribution of the classes is calculated in the final step using an output layer with sigmoid activation.

2) LSTM: The Long Short-Term Memory network (LSTM) is a Recurrent Neural Network (RNN) capable of learning long-term dependencies. In this work, the embedding layer has 40 embedding vector features with a max input length of 300 words. We applied 100 LSTM hidden units, default dropout 0, and finally, the sigmoid activation function.

3) CNN+LSTM: Fig. 3 shows the block model of our CNN-LSTM architecture. Here is an embedding layer following a 1D convolutional layer with 32 filters of size three and a 1D max-pool layer of pool size 2. We used the ReLU activation function to add non-linearity. After that LSTM model of 100 units is added with the sigmoid activation function.

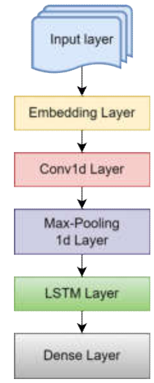

Fig. 3. Architecture of CNN+LSTM model

4) BiLSTM: Bidirectional LSTM (BiLSTM) is a sequence processing model made using two LSTMs, one for processing data in a forward direction and the other in a backward direction. The implementation of the BiLSTM model is quite similar to the LSTM model of this study, with the same parameter values.

5) CNN+BiLSTM: From Fig. 4 we see that CNN+BiLSTM also follows the CNN-LSTM architecture, with the only difference of adding the BiLSTM model instead of the LSTM model just after the convolutional layer.

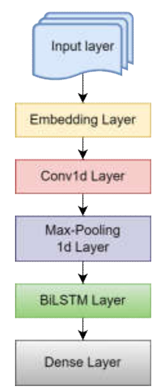

Fig. 4. Architecture of CNN+BiLSTM model

E. *Transformer Classifers*

1) mBERT [21]: The transformer model m-BERT is pre-trained over 104 languages taking about 110M parameters into account. We used BERT multilingual base model and finetuned it to apply to our news dataset. The batch size is 32 trained in five epochs.

2) Bangla-BERT [18]: Bangla-Bert is a pre-trained language model of the Bengali language using mask language modeling. We used the sagorsarker/Bangla-bert-base model and finetuned it to attain the best performance on our dataset. We set the batch size to 8 while training.

V. RESULT AND ANALYSIS

A. *Result on Final Dataset:*

We can see the performance of all the models of three approaches (ML, DNN, and Transformer) in Table II. It is evident that among all the Machine Learning approaches, SVM achieves better scores on Pr (95.2%), Re (95.2%), f1-score (95.2%), and Acc (95.2%) than all other models. LR is the closest to the SVM score, with an f1-score score (94.1%). MNB, KNN, DT, and Adaboost have an f1-score score of (91.5%), (88.5%), (85%), and (84.9%), respectively. DNN models have a significant improvement in terms of results. The CNN model has the best accuracy of 95.9% among all the models. But, CNN-LSTM and BiLSTM models are not far from the CNN model, with an accuracy of 95.5% and 95.3%, respectively. LSTM has 94.5% and CNN-BiLSTM offers 94.9% accuracy. Transformer models perform moderately on the fake news dataset. Bangla BERT model scores 94.1% accuracy, whereas mBERT performs 93.9%.

TABLE II. CLASSIFICATION RESULTS OF ALL MODELS

Method	Classifier	Precision	Recall	F1	Accuracy
Machine Learning	SVM	95.2	95.2	95.2	95.2
	LR	94.1	94.1	94.1	94.1
	MNB	91.6	91.5	91.1	91.5
	KNN	88.9	88.5	88.5	88.5
	DT	85	85	85	85
	AdaBoost	84.9	84.9	84.9	84.9
Deep Neural Network	CNN	**95.9**	**95.9**	**95.9**	**95.9**
	LSTM	94.5	94.5	94.4	94.4
	BiLSTM	95.3	95.3	95.3	95.3
	CNN+LSTM	95.5	95.5	95.5	95.5
	CNN+BiLSTM	94.9	94.9	94.9	94.9
Transformer model	Bangla-Bert-Base	94.1	94.1	94.1	93.3
	mBERT	93.9	93.9	93.9	93.8

TABLE III. MODELS PERFORMANCE ON SEPERATE DATASETS

Dataset	Criterion	Machine Learning						Deep Neural Network				
		MNB	*LR*	*KNN*	*SVM*	*DT*	*AdaBoost*	*CNN*	*LSTM*	*BiLSTM*	*CNN+LSTM*	*CNN+BiLSTM*
BanFakeNews [20] 55k real, 2k fake data (duplicate data exist)	Recall (Fake News classification)	46.6	47.8	35.9	87.6	93	92.5	94	91.5	99.4	90.3	92.9
	Average Accuracy	95.8	97.8	97	99.4	98.9	99	99.8	99.4	92.5	99.4	99.6
Bangla_Fake-Real_News Small_Dataset [6] 1.5k real, 1k fake data (duplicate data exist)	Recall (Fake News classification)	77.1	88.4	88.4	91.6	97	96.4	92.5	98.8	96.8	95.6	99.6
	Average Accuracy	91	95.4	95.3	96.8	98.1	97.5	96.2	97.5	97.5	97.6	88.5
Our Dataset 2.3k real, 2.3k fake (no duplicate data)	Recall (Fake News classification)	88.2	94.1	83.3	96	86	85.6	96.3	93.8	94.2	93.5	94.9
	Average Accuracy	94.1	88.5	95.2	85	84.9	94.4	95.9	95.3	95.9	94.9	93.3

B. *Recall and Accuracy Comparison on Seperate Datasets:*

We applied our designed fake news detection models to publicly available datasets separately. It helps us to understand the model efficiency compared with previous benchmark scores. From Table III, we can see that ML classifier models suffer badly in detecting fake news. That's why the recall value is low. But, some ML models, especially SVM, DT, and AdaBoost, can produce high accuracy on a massively imbalanced dataset. Thus, the accuracy does not represent actual fake news detection efficiency. However, on the BanFakeNews [20] dataset, DNN models perform well. BiLSTM model with 99.4% recalls on fake news and CNN with 99.8% highest accuracy. The Bangla Fake-Real News Small Dataset [6] has a similar pattern in terms of recall, with LSTM scoring 98.8% on fake news recall, whereas Decision Tree surprisingly has the best accuracy of 98.1%. The best performing model on the Final Dataset is BiLSTM, with an accuracy of 95.9%. CNN has a recall of 96.3% of fake news detection on that dataset. Due to computational limitations, we could not train our transformer models with the BanFakeNews [20] dataset, so unable to compare the outcomes from the transformer models

C. *Comparison of the Proposed Models with Previous Best Performing Studies*

The highest accuracy attained from BanFakeNews [20] data was 96% using GloVe and FastText on the top BiLSTM model. Our proposed LSTM model has a 3.8% improvement in accuracy on that dataset. The Bangla FakeReal Small Dataset showed 96.67% accuracy with the SVM model in [6], which is very consistent with our SVM model. However, in our experiment, Adaboost performs better on this dataset with an accuracy of 98.1%. So, we set a better benchmark for the dataset.

TABLE IV. COMPARISON WITH PREVIOUS TWO BENCHMARK SCORES

Dataset		**Classifier**	**Accuracy %**
BanFakeNews [20] 55k real, 2k fake data (duplicate data exist)	Previous work best performance	BiLSTM	96
	Our model performance	**LSTM**	**99.8**
Bangla_Fake-Real_News Small_Dataset [6] 1.5k real, 1k fake data (duplicate data exist)	Previous work best performance	SVM	96.7
	Our model performance	**AdaBoost**	**98.1**

D. *ROC curves and AUC of our proposed models*

A Receiver Operator Characteristic (ROC) curve is a graphical representation of True Positive Rate plotted against False Positive Rate and is used to show the diagnostic ability of binary classifiers. And the area underneath the ROC curve is called Area Under Curve (AUC). The closer the curve is to the top-left corner, the better the performance. Fig. 5 shows that our designed SVM and LR classifiers have the best ROC curves among ML algorithms. DNN classifiers have a very similar ROC pattern (Fig. 6). Bangla-BERT has an AUC score of 0.925 (Fig. 7).

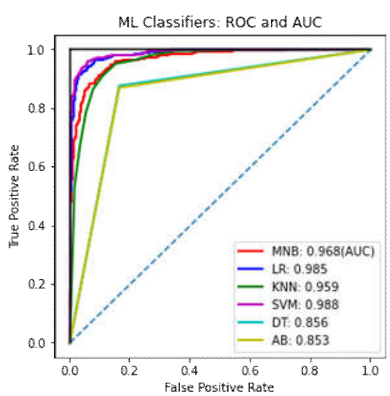

Fig. 5. ROC and AUC for ML Classifiers

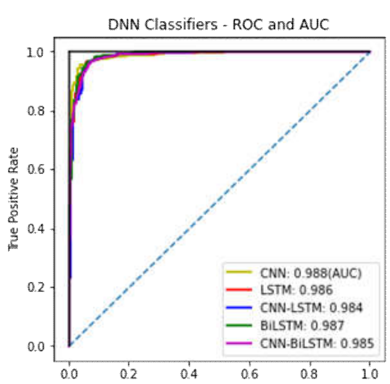

Fig. 6. ROC and AUC for DNN Classifers

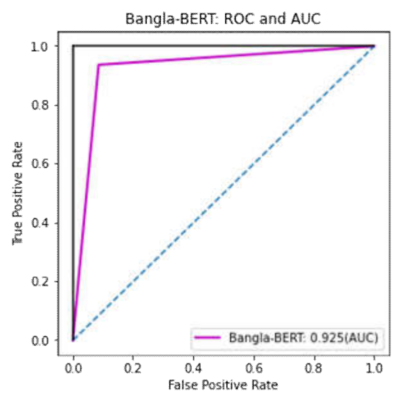

Fig. 7. ROC and AUC for Bangla-BERT

VI. CONCLUSION

This study combines prior research conducted by other researchers, then delivers a proper and better corpus of Bangla fake news and real news. Previous works of Bangla Fake news detection greatly suffered from poor data quality, like duplicate entries and imbalanced datasets. We remove all the redundant Fake news data and fix an equal ratio of fake and real news. Then we merge them with our collected Fake news data. We applied state-of-art machine learning, deep learning, and transformer models for benchmarking dataset performance. Among them, DNN models were the most remarkable ones with better performance scores. CNN model also showed the most promising results. Among the ML models, SVM performed exceptionally well, like previous works. The transformer models showed a good outcome as well. Our models produce a state-of-the-art result for the previously available fake news datasets with a 1.4% to 3.8% increase in accuracy and much better recall on detecting fake news. However, to improve our Bangla fake news dataset, we plan to add more fake news data from different sources that are different from the ones in the corpus.

REFERENCES

[1] "Buddhist temples, homes burned, looted in Ramu," bdnews24.com, 29 Sept 2012 [online]. Available: https://bdnews24.com/bangladesh/buddhist-temples-homes-burned-looted-in-ramu. [Accessed August 21, 2022].

[2] "4 die in Bhola as violence erupts over FB post," theindependent, 21 October, 2019 [online]. Available: https://m.theindependentbd.com/arcprint/details/220317/2019-10-21. [Accessed August 21, 2022].

[3] Star Report, "Mobs beat five dead for 'kidnapping'," The Daily Star, Sun Jul 21, 2019 [online]. Available: https://www.thedailystar.net/frontpage/news/mobs-beat-2-dead-kidnapping-1774471. [Accessed August 21, 2022].

[4] H. Ahmed, I. Traore, S. Saad, "Detecting opinion spams and fake news using text classification," Security and Privacy, 2018; 1:e9. https://doi.org/10.1001/spy2.9

[5] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi and B. -W. On, "Fake News Stance Detection Using Deep Learning Architecture (CNN-LSTM)," in *IEEE Access*, vol. 8, pp. 156695-156706, 2020, doi: 10.1109/ACCESS.2020.3019735.

[6] M. G. Hussain, M. Rashidul Hasan, M. Rahman, J. Protim and S. Al Hasan, "Detection of Bangla Fake News using MNB and SVM Classifier," *2020 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*, 2020, pp. 81-85, doi: 10.1109/iCCECE49321.2020.9231167.

[7] T. Islam, S. Latif and N. Ahmed, "Using Social Networks to Detect Malicious Bangla Text Content," *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, 2019, pp. 1-4, doi: 10.1109/ICASERT.2019.8934841.

[8] A. M. Ishmam and S. Sharmin, "Hateful Speech Detection in Public Facebook Pages for the Bengali Language," *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, 2019, pp. 555-560, doi: 10.1109/ICMLA.2019.00104.

[9] S. B. S. Mugdha, S. M. Ferdous and A. Fahmin, "Evaluating Machine Learning Algorithms For Bengali Fake News Detection," *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, 2020, pp. 1-6, doi: 10.1109/ICCIT51783.2020.9392662.

[10] F. Islam *et al*., "Bengali Fake News Detection," *2020 IEEE 10th International Conference on Intelligent Systems (IS)*, 2020, pp. 281-287, doi: 10.1109/IS48319.2020.9199931.

[11] T. Sraboni et al. "FakeDetect: Bangla fake news detection model based on different machine learning classifiers," B.Sc Dissertation, Department of Computer Science and Engineering, Brac University, Bangladesh, 2021, [online]. Available: http://hdl.handle.net/10361/14979.

[12] E. Hossain et al., "A Study Towards Bangla Fake News Detection Using Machine Learning and Deep Learning," AISC, vol 1408. Springer, Singapore. https://doi.org/10.1007/978-981-16-5157-1_7

[13] M. Z. H. George, N. Hossain, M. R. Bhuiyan, A. K. M. Masum and S. Abujar, "Bangla Fake News Detection Based On Multichannel Combined CNN-LSTM," *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2021, pp. 1-5, doi: 10.1109/ICCCNT51525.2021.9580035.

[14] A. S. Sharma, M. A. Mridul, and M. S. Islam, "Automatic Detection of Satire in Bangla Documents: A CNN Approach Based on Hybrid Feature Extraction Model," arXiv preprint arXiv:1911.11062, 2019.

[15] M. R. Hossain, M. M. Hoque, N. Siddique, I. H. Sarker, "Bengali text document categorization based on very deep convolution neural network," Expert Systems with Applications, Volume 184, 2021, 115394, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2021.115394.

[16] T. Alam, A. Khan, and F. Alam, 2020, "Bangla Text Classification using Transformers," arXiv preprint arXiv:2011.04446.

[17] T.T. Aurpa, R. Sadik, and M.S. Ahmed, "Abusive Bangla comments detection on Facebook using transformer-based deep learning models." Soc. Netw. Anal. Min. 12, 24 (2022). https://doi.org/10.1007/s13278-021-00852-x.

[18] M. Kowsher, A. A. Sami, N. J. Prottasha, M. S. Arefin, P. K. Dhar and T. Koshiba, "Bangla-BERT: Transformer-Based Efficient Model for Transfer Learning and Language Understanding," in *IEEE Access*, vol. 10, pp. 91855-91870, 2022, doi: 10.1109/ACCESS.2022.3197662.

[19] A. H. Zihad, 2022, "500 Bangla Fake News," [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/4222728

[20] Md Zobaer Hossain, Md Ashraful Rahman, Md Saiful Islam, and Sudipta Kar. 2020. "BanFakeNews: A Dataset for Detecting Fake News in Bangla," In Proceedings of the Twelfth Language Resources and Evaluation Conference, pages 2862–2871, Marseille, France. European Language Resources Association.

[21] J. Devlin et al., "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," DOI.org (Datacite), https://doi.org/10.48550/ARXIV.1810.04805.