2022 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE) | 979-8-3503-1156-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/WIECON-ECE57977.2022.10150461

FaND-X: Fake News Detection using Transformer-based Multilingual Masked Language Model

MD. Sijanur Rahman, Omar Sharif, Avishek Das, Sadia Afroze, Mohammed Moshiul Hoque*
[1]*Department of Computer Science and Engineering,*
Chittagong University of Engineering & Technology, Chattogram 4349, Bangladesh
{sijanurrahman2015, sadiacse10}@gmail.com {omar.sharif, avishek, moshiul_240}@cuet.ac.bd

***Abstract*—Recently, researchers have become seriously concerned about the rapid spread of false information or fake news and its harmful impact on societies. Although most fake news detection systems have been developed for English, it is a primitive stage concerning low-resource languages, especially Bengali. Fake news sometimes floods the virtual media, making it arduous to process and classify them manually. Thus, an automated system can be a good solution for detecting fake news with less time and cost. This paper introduces a framework (named *FaND-X*) for fake news detection by exploiting transformer-based and neural network-based approaches. This research also developed a dataset (called *BFNC*) consisting of 5K fake news data. In the experimentation, five machine learning (LR, RF, MNB, SVM, DT), three deep neural networks (CNN, BiLSTM, CNN+BiLSTM), and three transformer-based approaches (Bangla-BERT, m-BERT, XLM-R) were applied for the fake news classification task. Results show that the XLM-R beats all other techniques, as evidenced by the experiment's achieving a maximum f_1-score of 98% on the test data.**

***Keywords*—Natural language processing, Text processing, Fake news, Deep learning, Corpus**

I. Introduction

Fake news refers to a term used to describe articles that might inadvertently mislead or confuse readers by presenting false information [1]. The vast majority of fake news is created and distributed with malicious intent, which may harm a company's, organization's, or individual's reputation. Due to its accessibility and low cost, social media is already a common source of information sharing. However, because fake news, or news containing purposefully incorrect information, is widely spread on social media, getting news there has its drawbacks. Fake news production and dissemination have increased dramatically over the past few years to the point where it might be regarded as an infodemic. Groups, nation-states, and individuals spread such false information. To prevent rumours and misinformation from spreading, we need to have a system to detect fake news before spreading and causing heavy damage.

Due to the proliferation of social media usage and the lack of monitoring or filtering tools, the spreading rate of misinformation or fake news (in Bengali) has increased exponentially in recent years. It has been extensively studied to detect fake news in English and other high-resource languages. However, fake news detection is preliminary concerning low-resource languages, including Bengali. Thus, detecting fake news in Bengali is a crucial research concern among Bengali language processing experts. Developing an automatic system to detect fake news in Bengali is challenging due to the unavailability of standard datasets, lack of language processing tools, and complicated morphological structure of the Bengali language. News can be sarcastic or hilarious. When sarcasm and humour are heavily utilized, it would be difficult for even the most intelligent person to differentiate between authentic and fake news. Recently several approaches have been introduced to tackle the problem as far as the English Language is concerned. However, due to the language diversity and construct, a model developed in English cannot be used for Bengali. Thus, a new model must be developed and tuned based on the Bengali fake news dataset. This research intends to contribute to the following by taking into account the constraints of existing Bengali fake news detection:

- Create an annotated Bengali fake news dataset (called BFNC) consisting of 5K text data with 107088 unique words.
- Develop an automatic transformer-based framework (FaND-X) for detecting fake news from Bengali texts.
- Investigate the performance of various baseline machine learning (LR, RF, MNB, SVM, DT), deep learning (CNN, BiLSTM) and transformer-based techniques (Bangla-BERT, m-BERT, XLM-R) on the developed datasets to find the suitable method for the task.

II. Related Work

Researchers have been experimenting with various approaches to develop a rapid and automatic way to identify fake news in current years. A deep ensemble framework for multi-class classification has been proposed [2]. BiLSTM and CNN have been performed on the LIAR dataset with five classes. The best precision value it came up with is

0.88. A straightforward yet efficient method for tying fact-checking and fake news identification is proposed [3]. The strategy uses a text summarizing technique pre-trained on news corpora to condense the lengthy news story into a quick affirmation. In another work, various Russian-language false news detection models were trained and contrasted [4]. For this assignment, they used BERT embeddings, bag-of-n-grams features, a bag of Rhetorical Structure Theory, and other language features. On this task, comparing the model scores with the human scores suggested that the machine models were more adept at detecting fake news. A recent technique for creating altered (potentially fake) Arabic news reports is presented in [5]. An intelligent technique for fake news identification is created in [6]. They proposed a BERT+ALBERT model that achieved an accuracy of 0.855.

In Bengali, a study on fake news detection used linguistic features [7]. In a different study, they also added a sentiment as a feature to see how it would affect the process [8]. Trials were carried out using machine learning (ML) methods to identify Bengali fake news. Using an SVM, they achieved a maximum accuracy score of 73.20%. With CountVectorizer and TF-IDF as feature extraction, Multinomial Naive Bayes (MNB) and Support Vector Machine (SVM) classifiers have been applied to identify fake news in Bengali [9]. Their proposed framework determines whether an article is a piece of fake news based on its polarity. Finally, their analysis demonstrates that SVM with the linear kernel surpasses MNB with an accuracy of 96.64%. Several transformer-based models have been applied for fake news detection: M-BERT [10], RoBERTa [11], and BERT [12]. Concerning Bengali, most fake news detection systems show unreliable performance due to the small dataset and inappropriate usage of feature extraction and classification models. This research uses a more robust transformer-based model with better textual feature extraction capability that improves fake news detection accuracy.

III. BFNC: Bengali Fake News Corpus

The EU Commission defined *fake news* as disinformation in 2018 and defined it as demonstrably false or misleading that is manufactured, published, and disseminated with the intent to harm the public or gain financial gain [13]. Due to the lack of a standard corpus, we created a corpus (referred to as *BFNC* in the following) to identify fake news in Bengali text. This research will abide by the EU requirements because the activity relies on intentional false information produced to deceive people and sow social unrest.

Three human crawlers were tasked to acquire texts from various online sources. The news was crawled from well-known websites that offer news in Bengali. The websites' Alexa rankings have been taken into account. The news that the highest rankers provide is typically far more reliable. After collection, the corpus holds 5178 text data. This process costs almost eight months (January 2021 and August 2021). Table I shows an example of authentic and fake news from the BFNC dataset.

Table I
Data Examples

Data	Class
বেসরকারি পর্যায়ে পেনশন চালু করতে কাজ করছে সরকার: প্রধানমন্ত্রী (Government is working to introduce pension at private level: Prime Minister.)	Authentic
মুরগির হামলায় শেয়াল নিহত। তিন মুরগিকে গ্রেফতারের কথা জানিয়েছেন সংশ্লিষ্ট একজন। (Fox killed by chicken attack. A person concerned informed about the arrest of three chickens.)	Fake

The accumulated data is given to three postgraduate students for annotations. Before annotation, examples of each category were provided to the annotators. Annotators assign a label from 0 or 1 to a text. The fake and authentic classes are represented here by 0 and 1. A majority vote is used to determine a text's label [14]. As the initial label for the text, the label that received the most votes from the annotators is chosen.

Each labelling performed by the annotators was personally validated by a specialist who has been an academician working on NLP for more than 20 years. If the initial label supplied by the annotators agrees with the expert's initial label, it is considered the final. An algorithm proposed by Iqbal et al. [15] was used to perform cross-validation. Finally, the BFNC contained confirmed 5048 data (where 3049 are authentic and 1999 are fake news) with their appropriate labels and stored in *.xlsx format. We achieved a Cohen's Kappa of $\boldsymbol{K = 0.90}$, indicating that the inter-annotator reliability is good [16].

To assess the models, the data were separated into a train set (4039 texts), a validation set (503 texts), and a test set (506 texts). Table II shows several statistics of the training set.

Table II
Statistics of the train set.

Dataset	Train	Validation	Test	Words	Unique Words
Authentic	2440	304	305	506462	56299
Fake	1599	199	201	261503	50789
Total	4039	503	506	767965	107088

IV. Methodology

This study aims to develop an automatic framework that recognizes fake news from Bengali text. Various ML, DL and transformer-based techniques are exploited with respective feature extraction (TF-IDF) and embedding techniques (Word2Vec, Fast-Text and GloVe). All models are trained and fine-tuned using the BFNC.

- **TF-IDF**: This technique assesses the significance of a word to a record within a group of documents. The 30000 words in the corpus that occur most frequently are used to extract unigram, bigram and trigram features.
- **Word2vec**: It was trained on Skip-gram and a window dimension of 5. We considered a minimal word count of 3 and 600 embedding dimensions.
- **FastText**: During training, if a term is not present in the lexicon, it can be created from its component n-grams. The models were trained on Skip-gram with a character n-gram length (5). A window length of 5 and 600 embedding dimensions are considered.
- **GloVe**: Bengali language vectors that have already been trained are available for GloVe and deployed here. The embedding dimension utilized here is 300.

A. Classifiers

Five ML, three DL and three transformer-based models are explored as the classifier models.

- **ML Approaches**: Using a TF-IDF text vectorizer, five ML techniques (LR, SVM, RF, DT and MNB) are applied. The "lbfgs" solver for LR is selected, along with the "l2" penalty, and the C value is settled to 1. SVM uses a C value of 1 with the "rbf" kernel. The n-estimators parameter for RF is set to 300, the min-samples-leaf parameter to 30, and the MNB parameter is set to alpha = 1.0. There was a 400 iteration cap. Table III provides a summary of the ML parameters.

Table III
Optimized parameters for ML models

Classifier	Optimal Parameters
LR	max-iter = 400, C=1, penalty='l2', solver ='lbfgs'
RF	n-estimators=300,criterion='gini',max-features='sqrt',min-samples-leaf=30,random-state=100
MNB	alpha = 1.0, fit-prior = False, class-prior = None
SVM	kernel='rbf', random-state = 0, gamma='scale', tol='0.001'
DT	max-depth=3, random-state=42

- **CNN:** The Word2Vec, FastText, or Glove embeddings' training weights are sent to the embedding layer, which creates a sequence matrix. A two-layer CNN architecture propagates embedding features. Each convolutional layer has one hundred twenty-eight neurons, and the kernel size is (1×2). Convolution layer output is max-pooled over time and sent to a layer with complete connectivity. The associated layers employ "ReLU" activation. The final hidden representations are created by this FC layer, which has two neurons, using the pooling characteristics. One output layer is designed to estimate the class probability using the softmax activation function.
- **BiLSTM:** The attention mechanism is an effective pooling technique for classification tasks when used with Bi-LSTM. An embedding layer resembling CNN is a component of the newly designed BiLSTM network. One BiLSTM layer with 128 units composes the built-in network. The output of this layer is pushed to a fully connected layer consisting of two neurons. The classification is then carried out using a softmax function. A dropout layer (with a rate of 20%) is applied before the softmax operation.
- **CNN+BiLSTM:** An embedding layer, two 1D convolutional layers with 128 neurons, each with a kernel size of (1×2), and a BiLSTM layer with 128 units make up this combined network. A 1D max-pool layer follows the BiLSTM layer. One output layer is designed to estimate the class probability using the softmax activation function.

Table IV represents a summary of the deep learning models' hyperparameters.

B. Transformer Models

A transformer model called Bidirectional Encoder Representations from Transformers (henceforth, BERT) [17] was created to pre-train deep bidirectional representations from unobserved data. We applied three transformer models to our dataset: Bangla-BERT, m-BERT, and XLM-R. The models are selected from the Huggingface transformers library and refined using the Ktrain program on the BFNC corpus.

- **Bangla-BERT:** A pre-trained BERT mask language modeling called Bangla-BERT [18] was developed using a sizable Bengali corpus. We updated the pre-trained model fitted for our created corpus using the "sagorsarker/Bangla-bertbase" model, which we then fine-tuned. To provide a better performance, a batch size of 16 is used.
- **m-BERT:** A pre-trained transformer model with more than 110M parameters is called m-BERT [17]. With a batch size of 12, we used the "bert-base-multilingual-cased" model to refine it on BFaNC.

V. FaND-X: Proposed Fake News Detection Model

Figure 1) outlined the suggested architecture for fake news detection that uses the transformer-based Multilingual Masked Language Model (XLM-R) [19].

The XLM-R network receives labelled train data after cleaning and preprocessing. This approach is based on extensive cross-lingual unsupervised learning. This research used the xlm-roberta-base, a scaled-down version of the model with 270M parameters, 12 layers, 768 hidden states, and eight heads. This model was trained using 2.5 TB of freshly generated, clean Common Crawl data in 100 languages. The network's fully - connected layers are replaced with a set of output layers during the fine-tuning phase. The pre-trained model is trained using the

Table IV
Optimized hyperparameters for deep learning models

Hyperparameters	Hyperparameter Space	CNN	BiLSTM	CNN+BiLSTM
Filter Size	3,5,7,9	3	-	7
Pooling Type	'max', 'average'	'max'	-	'max'
Embedding Dimension	30, 35, 50, 70, 90, 100, 150, 200,250, 300	300	300	300
Number of Units	16, 32, 64, 128, 256	128	128	128
Neurons in Dense Layer	16, 32, 64, 128, 256	-	16	-
Batch Size	16, 32, 64, 128, 256	32	32	32
Activation Function	'relu', 'tanh', 'softplus', 'sigmoid'	'relu'	'relu'	'relu'
Optimizer	'RMSprop', 'Adam', 'SGD', 'Adamax'	"Adam"	"Adam"	"Adam"
Learning Rate	0.5, 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005, 0.0001	0.001	0.001	0.001

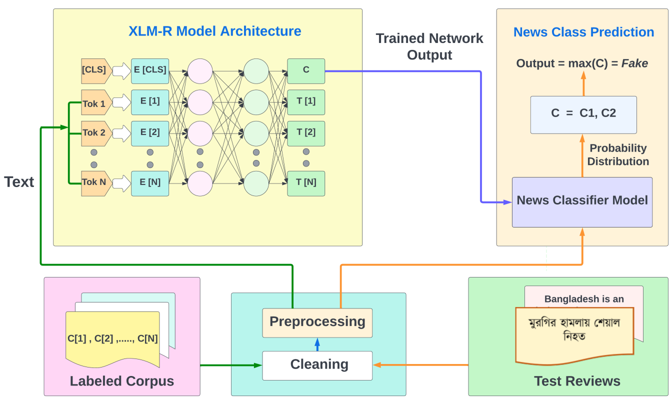

Figure 1. Overall architecture of the FaND-X model for fake news detection

provided dataset in this method. The news classification model receives the output from the trained neural network. After performing probability distribution on the test data set, this trained classifier model predicts the news class. We applied the "xlm-Roberta-base" model to our generated BFaNC corpus with a batch size of 16.

Each hyperparameter's initial value is chosen randomly except for the max sequence length. We go over the hyperparameter space to determine a hyperparameter's ideal value. All of the transformer models underwent a 25-epoch training process. The best intermediate model is retained and used to forecast the test data at the checkpoint. Table V shows the list of initial and optimal parameters of transformer models.

Table V
Optimized hyperparameters for transformers models

Hyperparameter	Initial Value	Optimal Value (BFaNC)	Optimal Value (English)
Fit method	'auto-fit'	'auto-fit'	'auto-fit'
Learning rate	1e-5	2e-5	2e-5
Epochs	25	20	20
Batch size	4	16	12
Max sequence length	-	256	50

VI. Results and Analysis

The weighted *f1-score* is employed to determine the better-performing model. However, the criteria for accuracy, recall, and precision are also considered. Table VI shows the evaluation results for each model.

Among ML approaches, LR achieved the highest (89%) f_1-*score* than RF (86%), MNB (86%), SVM (86%) and DT (86%). In DNNs, CNN+BiLSTM with GloVe embedding technique outperformed other approaches as far as all the evaluation parameters are concerned. It attained f_1-*score* of (91%). However, CNN+BiLSTM (GloVe) acquired an approximately 2.1% higher f_1-*score* than the best ML model (i.e., LR).

Mentionable improvement in all scores is observed utilizing transformer-based approaches. As far as transformer-based models are concerned, m-BERT achieved the lowest of 84% f_1-*score*. Additionally, Bangla-BERT shows a 4% improved f_1-*score* (88%) than m-BERT (84%). XLM-R model delivers an enormous headway of nearly 14% compared to m-BERT and 10% corresponded to Bangla-BERT, respectively. It attained an f_1-*score* of 98%, which is the most elevated of all the performed models. It is observed from Table VI that LR performed the best among ML models. CNN + BiLSTM (GloVe) performed the best among DNN models. However, the proposed XLM-R model outperformed all other models. The calibre of the training corpus strongly influences any model's outcome. Here, the experiments are carried out on the developed corpus BFNC. The superior quality of the corpus ensured a competitive result. Another considerable factor is the method. Here, our experiment exploits the cross-lingual property of the XLM-R model, which was trained in 100 different languages. Combining the quality of the corpus with the cross-lingual property of XLM-R paved the way for high performance. Table VII projects the class-wise performance of the proposed model.

Table VI
Comparison of various approaches on test set

Method	Classifier	Accuracy	Precision	Recall	f_1-score
ML models	**LR+TF-IDF**	**0.89**	**0.89**	**0.89**	**0.89**
	RF+TF-IDF	0.87	0.87	0.87	0.86
	MNB+TF-IDF	0.87	0.87	0.87	0.86
	SVM+TF-IDF	0.87	0.87	0.87	0.86
	DT+TF-IDF	0.87	0.87	0.87	0.86
DNN models	CNN(One Hot Encoding)	0.87	0.86	0.86	0.86
	CNN(Word2Vec)	0.78	0.77	0.76	0.76
	CNN(FastText)	0.83	0.83	0.83	0.83
	CNN(GloVe)	0.90	0.90	0.90	0.90
	BiLSTM(One Hot Encoding)	0.90	0.90	0.90	0.90
	BiLSTM(Word2Vec)	0.80	0.82	0.79	0.79
	BiLSTM(FastText)	0.86	0.85	0.86	0.85
	BiLSTM(GloVe)	0.89	0.89	0.89	0.89
	CNN + BiLSTM(One Hot Encoding)	0.90	0.89	0.90	0.89
	CNN + BiLSTM(Word2Vec)	0.81	0.80	0.81	0.80
	CNN + BiLSTM(FastText)	0.84	0.84	0.84	0.84
	CNN + BiLSTM(GloVe)	**0.91**	**0.91**	**0.91**	**0.91**
Transformers	Bangla-BERT	0.88	0.88	0.88	0.88
	m-BERT	0.84	0.84	0.84	0.84
	XLM-R(Proposed)	**0.98**	**0.98**	**0.98**	**0.98**

Table VII
Class-wise performance of best ML, DNN and proposed models on the test set

Approach	Class	Precision	Recall	f_1-score
Monolingual LR	Authentic	0.89	0.94	0.91
	Fake	0.89	0.82	0.85
Monolingual CNN+BiLSTM (GloVe)	Authentic	0.90	0.95	0.93
	Fake	0.92	0.83	0.88
Multilingual **XLM-R(Proposed)**	Authentic	**0.97**	**0.98**	**0.97**
	Fake	**0.99**	**0.98**	**0.98**

A. Error Analysis

Table VI makes it evident that XLM-R is most helpful in identifying fake news. Using the confusion matrix, this research performed an exhaustive error analysis. Figure 2 shows a class-wise portion of the total number of anticipated labels.

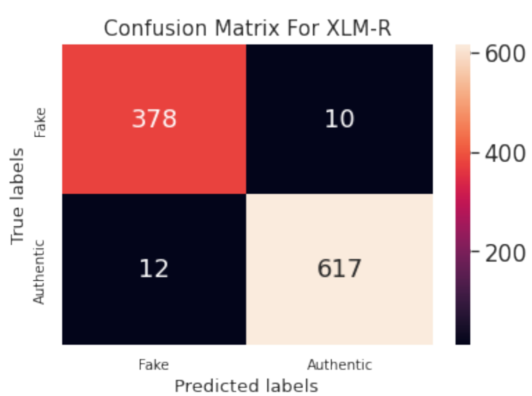

Figure 2. Confusion matrix of XLM-R model

It is obeyed that few data are classified wrongly. Ten news texts are fake. However, our system detected them to be authentic news. 12 authentic news was predicted as fake news. The error analysis demonstrates that fake news attained a higher rate of accurate classification (98.9%) while authentic news accumulated (96.7%). There was some authentic news that our model predicted to be fake news. A critical feature of authentic news is the specific mention of date and time. As those authentic pieces of news did not mention the date and time for some reason, our system mispredicted them as fake news.

B. Comparison with Existing Works

The outcomes of the offered approach (XLM-R) are compared with the existing techniques ([20], [7], [21], [22]) to assess its effectiveness. Table VIII shows the upshots of the comparison. The results suggested that the XLM-R model surpassed earlier approaches by obtaining the greatest f_1-*score* of **(98%)** for detecting fake news in Bengali.

Table VIII
Results of comparison

Method	*f1-score*
AraBERT [20]	0.67
L+POS+E(F) [7]	0.92
Hybrid CNNs [21]	0.83
RoBERTa [22]	0.71
XLM-R (Proposed)	0.98

VII. Conclusion

This research examined several ML, DL, and transformer-based methods to identify fake news in Bengali. The experimental analysis revealed that the XLM-R model delivered the most elevated results on the developed dataset (BFNC) detecting fake news in Bengali among all explored approaches. Specifically, XLM-R obtained the highest f_1-score of 98%, indicating an improvement of 9% (over ML) and 7% (compared to DNN). Moreover, the XLM-R model also outperformed the current works. Due to the proliferation of journalistic cliches and extensive usage of satire and humour, the proposed model misclassified some texts in the test set. Although the XLM-R demonstrated the highest score, alternative approaches (for example, transformer-based ensembling) can also be investigated for enhanced outcomes. Classification of fake news in false context, satire, and clickbait after detection can be performed. Additionally, the developed BFNC may be extended into a broader range, and more robust transformer-based models can explore in future to improve performance.

References

[1] M. M. H. Omar Sharif, "Tackling cyber-aggression: Identification and fine-grained categorization of aggressive texts on social media using weighted ensemble of transformers," *Neurocomputing*, vol. 490, pp. 462–481, 2022.

[2] A. Roy, K. Basak, A. Ekbal, and P. Bhattacharyya, "A deep ensemble framework for fake news detection and multi-class classification of short political statements," in *Proceedings of the 16th International Conference on Natural Language Processing*, 2019, pp. 9–17.

[3] Q. Li and W. Zhou, "Connecting the dots between fact verification and fake news detection," *arXiv preprint arXiv:2010.05202*, 2020.

[4] G. Kuzmin, D. Larionov, D. Pisarevskaya, and I. Smirnov, "Fake news detection for the russian language," in *Proceedings of the 3rd International Workshop on Rumours and Deception in Social Media (RDSM)*, 2020, pp. 45–57.

[5] A. Elmadany, M. Abdul-Mageed, T. Alhindi *et al.*, "Machine generation and detection of arabic manipulated and fake news," in *Proceedings of the Fifth Arabic Natural Language Processing Workshop*, 2020, pp. 69–84.

[6] R. Vijjali, P. Potluri, S. Kumar, and S. Teki, "Two stage transformer model for covid-19 fake news detection and fact checking," *arXiv preprint arXiv:2011.13253*, 2020.

[7] M. Z. Hossain, M. A. Rahman, M. S. Islam, and S. Kar, "Banfakenews: A dataset for detecting fake news in bangla," *arXiv preprint arXiv:2004.08789*, 2020.

[8] M. Y. Tohabar, N. Nasrah, and A. M. Samir, "Bengali fake news detection using machine learning and effectiveness of sentiment as a feature," in *2021 Joint 10th International Conference on Informatics, Electronics & Vision (ICIEV) and 2021 5th International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*. IEEE, 2021, pp. 1–8.

[9] M. G. Hussain, M. R. Hasan, M. Rahman, J. Protim, and S. A. Hasan, "Detection of bangla fake news using mnb and svm classifier," *arXiv preprint arXiv:2005.14627*, 2020.

[10] T. Pires, E. Schlinger, and D. Garrette, "How multilingual is multilingual bert?" *arXiv preprint arXiv:1906.01502*, 2019.

[11] D. Dementieva and A. Panchenko, "Cross-lingual evidence improves monolingual fake news detection," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: Student Research Workshop*, 2021, pp. 310–320.

[12] R. Pranesh, M. Farokhnejad, A. Shekhar, and G. Vargas-Solar, "Cmta: Covid-19 misinformation multilingual analysis on twitter," in *59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: Student Research Workshop*. Association for Computational Linguistics, 2021, pp. 270–283.

[13] A. Pathak and R. K. Srihari, "Breaking! presenting fake news corpus for automated fact checking," in *Proceedings of the 57th annual meeting of the association for computational linguistics: student research workshop*, 2019, pp. 357–362.

[14] D. Magatti, S. Calegari, D. Ciucci, and F. Stella, "Automatic labeling of topics," in *2009 Ninth International Conference on Intelligent Systems Design and Applications*. IEEE, 2009, pp. 1227–1232.

[15] M. Iqbal, A. Das, O. Sharif, M. M. Hoque, and I. H. Sarker, "Bemoc: A corpus for identifying emotion in bengali texts," *SN Computer Science*, vol. 3, no. 2, pp. 1–17, 2022.

[16] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[18] S. Sarker, "Banglabert: Bengali mask language model for bengali language understanding," 2021.

[19] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[20] E. M. B. Nagoudi, A. Elmadany, M. Abdul-Mageed, T. Alhindi, and H. Cavusoglu, "Machine generation and detection of arabic manipulated and fake news," *arXiv preprint arXiv:2011.03092*, 2020.

[21] J. C. B. Cruz, J. A. Tan, and C. Cheng, "Localization of fake news detection via multitask transfer learning," *arXiv preprint arXiv:1910.09295*, 2019.

[22] V. Slovikovskaya, "Transfer learning from transformers to fake news challenge stance detection (fnc-1) task," *arXiv preprint arXiv:1910.14353*, 2019.