IJCDS 1571057398

International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon20..)

http://dx.doi.org/10.12785/ijcds/XXXXXX

A Comprehensive Dataset and Deep Learning Approach for Misinformation Detection on Social Media in Bangladesh

Mohammad Rifat Ahmmad Rashid[1], Rahul Roy[2], Din M Sumon Rahman[2], Musa Akram Saleh[1], Abdul Ali Hayder Khan[1], Md. Abu Rayhan[1], Khandaker Foysal Ahmed[1], Nafees Monsoor[1] and Mahamudul Hasan[1]

[1]*Department of Computer Science and Engineering,East West University, Dhaka, Bangladesh*
[2]*Department of Computer Science and Engineering, University of Liberal Arts, Dhaka, Bangladesh*

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

Abstract:
In an effort to address the growing issue of misinformation on social media, particularly in the context of the Covid-19 pandemic, we have diligently developed a comprehensive dataset on Bangla misinformation. This dataset was scraped from FactWatch, a leading fact-checking organization in Bangladesh, and annotated with fact ratings. It includes a meticulously curated collection of 1014 fact-checked reports spanning from October 4, 2021, to May 25, 2023. These reports encompass a diverse array of summaries, categories, and reliable correctness labels, providing samples of the original fake news content along with investigative descriptions of the fact-checking processes employed. The dataset represents a significant contribution to Bangladesh's participation in the global effort to combat fake news and serves as a crucial resource for ongoing research in misinformation studies, natural language processing, and automated fact-checking, particularly for content in the Bengali language. Addressing the issue of misinformation within the under-researched Bangla language context, our study also leveraged this dataset for deep learning analysis, employing advanced techniques such as Long Short-Term Memory (LSTM) networks and Bidirectional Encoder Representations from Transformers (BERT) with a Bangla base model. The BERT model, with its robust Transformer architecture, excelled in linguistic analysis, achieving an accuracy of 98.77%, while the LSTM model, adept at handling sequential data, recorded an accuracy of 88.92%. The Bangla BERT base model demonstrated exceptional performance in precision, recall, and F1-score, marking a substantial advancement in misinformation detection for the Bangla language.
Keywords: Misinformation, Fact-Checking, Social Media Analysis, Natural Language Processing, Long Short-Term Memory (LSTM), Bidirectional Encoder Representations from Transformers (BERT), Data Annotation, Digital Information Reliability.

1. INTRODUCTION

In the digital age, the explosion of online content accessible through the internet and electronic devices has transformed the way we access information. Platforms such as social media, blogs, and websites have become integral to our daily lives, providing a wealth of information. However, this digital proliferation has a downside: the rampant spread of fake news, posing a significant challenge in maintaining the integrity of online content. This issue becomes even more acute in languages that lack substantial digital resources, such as Bangla. Despite its vast number of speakers, Bangla has not received sufficient technological support for detecting and combating misinformation. The problem of fake news is not new, but its impact is particularly severe in Bangla-speaking regions, which have not benefited as much from the technological advancements aimed at curbing misinformation prevalent in English-speaking areas. In these regions, digital platforms are crucial for spreading information but are also hotbeds for the unchecked circulation of false news. The scarcity of effective misinformation detection tools in the Bangla language not only hampers the fight against fake news but also leads to significant social and cultural issues within Bengali-speaking communities. The technological disparity in supporting diverse languages highlights a significant challenge in maintaining information reliability across various linguistic contexts.

Deep learning models, such as Long Short-Term Memory (LSTM)[1] and Bidirectional Encoder Representations from Transformers (BERT)[2], have proven to be highly effective in understanding and processing complex language structures. These models are pivotal for accurately identifying misinformation in text-based content. Addressing the specific challenge of detecting fake news in the Bengali language, various methodologies have been explored, primarily focusing on natural language processing (NLP) and machine learning (ML) techniques. Research efforts, as indicated in studies [3][4], have delved into ensemble methods and deep learning architectures like Bi-LSTM-

E-mail address: author 1, author 2, author 3

GRU-Dense, along with sentiment analysis, to improve detection accuracy, particularly in the realm of social media. There's also a recognized need for specialized approaches for low-resource languages, as highlighted in references, and for addressing the linguistic complexities in different contexts, such as Arabic. Efforts to identify fake social media accounts have employed advanced algorithms and data preprocessing, though these often don't fully address the unique challenges posed by Bengali language content on platforms like Facebook. Our proposed methodology integrates LSTM with the Bangla BERT base model, custom-designed for the Bengali language. This integration aims to fill the existing gap and offers a more effective tool for combating misinformation in this linguistic landscape.

In our study, we tackle the issue of fake news within the Bengali-speaking community using cutting-edge deep learning methods. We have compiled a dataset from a fact-checking website, tailored it for the Bangla BERT base model, and employed LSTM networks along with Bangla BERT for classification purposes. Our goal is to create a reliable tool for verifying Bengali news content, fostering a culture of fact-checking, and enhancing the accuracy of news reporting. Our process involves gathering and preparing Bengali news data, identifying important features, and implementing deep learning models, specifically LSTM and Bangla BERT, for categorization. Our dataset, comprising 2024 news items sourced from Fact Watch, underwent a rigorous preprocessing regimen, including labeling, cleaning, and segmentation into training, validation, and testing subsets. It is pre-trained on a comprehensive Bangla text corpus, underwent additional fine-tuning to specialize in fake news classification. To ensure the model's dependability, we implemented 5-fold cross-validation and meticulously adjusted hyperparameters to hone the model's performance.

This paper is structured as follows: The Literature Review explores existing research on fact validation, with an emphasis on identifying the research gap for the Bengali language. The Methodology section details the dataset acquisition, its preprocessing, and the implementation of LSTM and Bangla BERT models. The Results and Discussion section presents an experimental analysis of the dataset and the performance measures, showcasing the models' capabilities in detecting fake news in Bengali. Finally, the Conclusion and Future Work section encapsulates the principal discoveries and proposes potential avenues for further investigation in this research area.

2. LITERATURE REVIEW

The increasing prevalence of fake news in Bengali, amplified by the digital era, has led to diverse methodological approaches, especially in NLP, for its detection. The study [3] positioned fake news as a democratic hazard and underscored the significance of NLP and Machine Learning (ML) in classifying news articles while highlighting the challenges posed by the scarcity of robust datasets and specialized methods. It demonstrated promising outcomes in fake news identification using three ML classifiers. Similarly, [5] introduced an innovative neural network architecture aimed at enhancing the correlation between news content and factual events, achieving a notable accuracy of 94.21%. The work in [6] ventured into the realm of online information across various platforms, implementing an ensemble approach with a Bi-LSTM-GRU-dense model using the LIAR dataset. This approach not only achieved remarkable accuracy, recall, precision, and F-score metrics but also set new benchmarks for misinformation classification, advancing the research landscape in combating misinformation within Bengali-speaking communities.

The research detailed in [7] delved into the escalating issue of fake news within Bengali-speaking communities, employing machine learning techniques for its detection. While it integrated sentiment analysis into the detection framework, the study revealed that this approach did not substantially enhance the outcomes, with the most notable result being a 73.20% accuracy rate achieved through the use of a support vector machine. In contrast, [4] focused on the detection of online news, particularly emphasizing the utilization of sophisticated deep learning methods such as Attention mechanisms, GANs, and BERT. In a study from Thailand [8], researchers developed a comprehensive two-phase framework to tackle fake news, leveraging the synergy of information retrieval, NLP, and machine learning techniques. Initially, the framework employed a web-crawler for data collection from Thai news websites, followed by NLP-based feature extraction. Subsequent testing of various machine learning models identified the LSTM model as the most effective, leading to the creation of an automated application specifically for fact validation. The study in [9] addressed the escalating problem of misinformation in news, particularly magnified by social media, by pioneering a hybrid model that integrates convolutional and RNNs.

The study [10] introduced "FakeBERT," a novel deep learning model that combines BERT with a deep Convolutional Neural Network, adept at understanding semantic subtleties, achieving a remarkable 98.90% accuracy in fake news classification. Meanwhile, research [11] focused on Bengali content, utilizing a Gaussian Naive Bayes algorithm paired with TF-IDF for text analysis and an Extra Tree Classifier for attribute selection, effectively distinguishing between fake and real news with an 87% accuracy rate, showcasing its potential in the complex domain of online news detection.

In the realm of combating misinformation on social media, particularly during events with national security implications, the study [12] delves into a variety of detection methodologies, tools, and browser extensions, highlighting the pivotal role of feature extraction through Machine Learning and NLP in preserving the integrity of online information. The research in [13] critiques the reliance solely on NLP for fake news detection, introducing a system that integrates secondary features like source domains and author names, demonstrating improved accuracy over traditional NLP models. Concurrently, [14] pioneers the detection of misinformation in Bangla from social media, a relatively unexplored area, by employing SVM and Multinomial Naïve Bayes algorithms alongside CountVectorizer and TF-IDF Vectorizer for feature extraction, noting a superior performance with SVM and thereby advancing the field of fact validation in less commonly addressed languages.

Study [15] tackled the lack of annotated datasets in languages like Bangla, creating an extensive corpus of 50,000 news items and merging advanced NLP techniques with neural network-based methods to enhance automated detection systems. Research in [16] marked a pivotal shift towards automated methods by leveraging deep learning and neural network models, trained on textual data from both full articles and titles. Study [17] highlighted the effectiveness of word pairs in verifying news authenticity, demonstrating that random forest classifiers using bigram features achieved the highest accuracy in detecting online falsehoods through semantic features. Study [18] addressed the urgent need for misinformation detection during the COVID-19 ”infodemic,” evaluating models such as LSTM, GRU, and BiLSTM in differentiating truth from falsehood in English and Chinese texts, with BiLSTM showing up to 99% accuracy. Study [19] focused on Arabic fake news detection, employing LSTM and CNN models to analyze article-headline pairs, especially in the context of Middle Eastern politics, and showing promising results.

Study [20] introduced a method for identifying fake Twitter accounts using a stacked ensemble of machine learning algorithms, achieving a high accuracy rate of 99%. Study [21] presented a Twitter fake account detection system that used a BiGRU model combined with GloVe word embedding for tweet content analysis, outperforming standard models with an accuracy of 99.44% and precision of 99.25%. In response to the challenge of fake news impeding Thailand’s COVID-19 recovery, the paper [22] proposes an innovative Thai-NLP based approach using transfer learning models to detect misinformation in Thai texts. The method involves pre-training on globally sourced COVID-19 datasets, translated from English to Thai, and employs ’feature shifting’ to enhance Thai text representation, addressing the scarcity of local datasets.

In the paper [23] addressed the significant challenge of fake news, identified as a major cyber-security threat to Thai COVID-19 recovery efforts, by introducing a state-of-the-art detection system using Thai-NLP and transfer learning models. It employed innovative ’feature shifting’ and machine translation techniques to enhance Thai text datasets for model training, promising substantial improvements in the performance of misinformation detection during the fine-tuning stage. The paper [24] tackled the obstruction of fact validation in Thai COVID-19 recovery by proposing an advanced detection system using Thai-NLP and transfer learning, leveraging ’feature shifting’ and machine translation to significantly enhance model training and detection accuracy. The article [25] introduced Íntegro, a defense system that tackled the detection of counterfeit accounts on social media by innovatively classifying and ranking user accounts, leveraging victim classification.

Previous studies have predominantly focused on techniques like SVM, Naive Bayes, and Decision Trees, alongside the application of word embeddings and ensemble methods for enhanced accuracy. The advent of sophisticated models like Bi-LSTM-GRU-Dense and the utilization of large datasets, such as the LIAR dataset, have significantly advanced the field. However, these methods often lack the specific focus and adaptability required for Bengali language content on social media platforms. Our proposed method diverges from these approaches by integrating LSTM and the Bangla BERT base model, specifically fine-tuned for the intricacies of the Bengali language and the unique characteristics of Facebook content.

3. METHODOLOGY

Figure 1 illustrates the workflow diagram of our proposed approach that outline a multi-stage process. This workflow begins with Dataset Collection, where we meticulously source data from Fact Watch, a repository rich with news articles marked for credibility, and then convert this data into a structured, processable format. A crucial part of this stage involves preprocessing, which includes tasks such as noise removal, handling missing values, normalizing text, and converting it into a format suitable for deep learning models, ensuring that the data is primed for complex analysis. Advancing to the Model Training and Validation Process, our methodology involves setting up a robust deep learning framework. Here, we employ LSTM networks and BERT for training the model, carefully defining the architecture, selecting hyperparameters, and preparing for the training and validation tasks. The model receives new data inputs and applies its learned patterns, predicting the veracity of news content through an API-facilitated inference process. The workflow of the proposed approach consists of the following stages:

Dataset Collection: The initial stage involves sourcing the dataset from Fact Watch, which is likely a database containing various news articles marked for credibility. The sourced data is then converted into an appropriate import format, which prepares it for processing. This step ensures that the data is in a structured form that can be easily manipulated and understood by the subsequent processing algorithms. The raw data undergoes preprocessing where it’s cleaned and transformed. This may include removing noise, handling missing values, normalizing text, and perhaps converting it into a format suitable for training deep learning models. As a result of preprocessing, the data is now in a processed state, ready to be used for training and

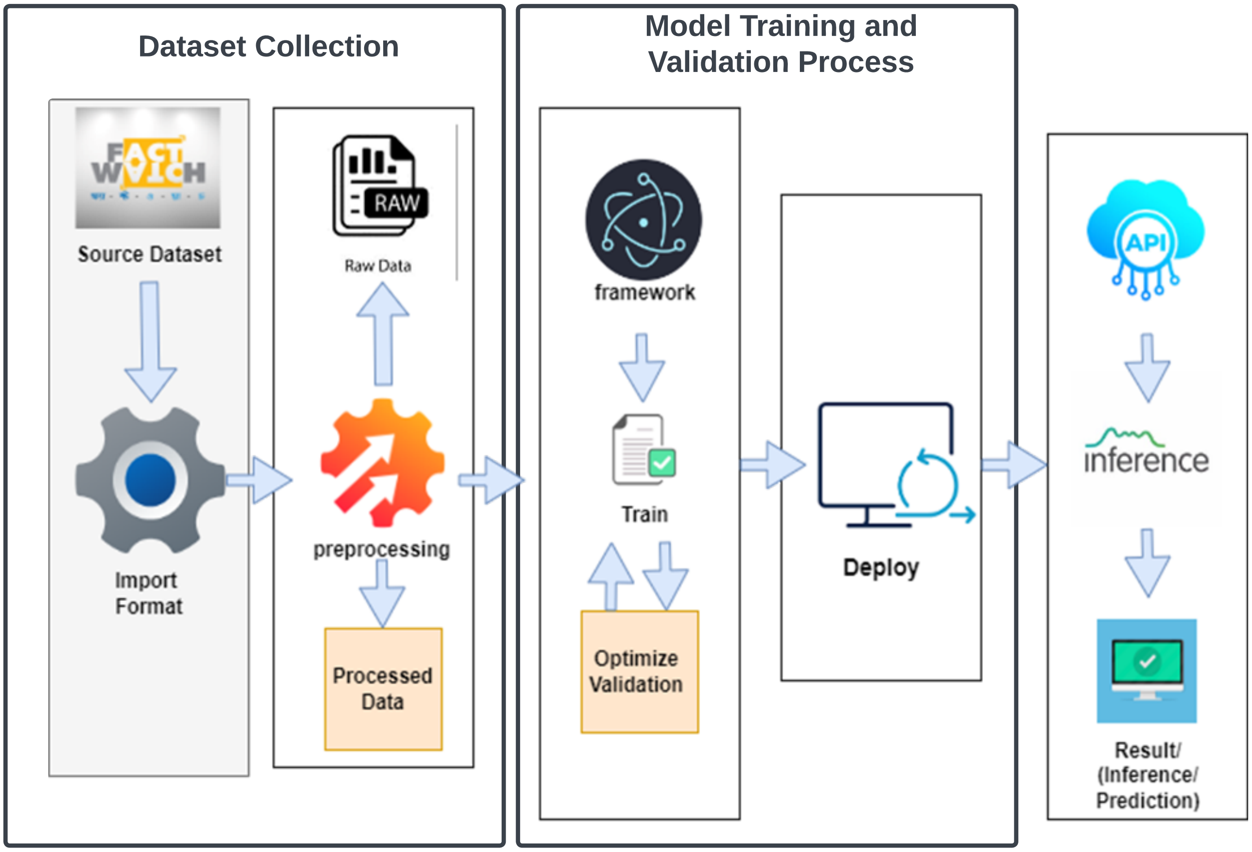

Figure 1. Proposed Workflow diagram.

validating tasks.

Model Training and Validation Process: This step involves setting up the deep learning framework, which includes defining the model architecture, selecting hyper-parameters, and preparing for training and validation tasks. This involves the model learning from the data's features to make accurate predictions about the news articles' authenticity. After training, the model is validated and optimized. This step likely involves tuning the model with validation data to improve its accuracy and reduce overfitting or underfitting. Once the model is trained and validated, it's deployed to a production environment where it can start making inferences or predictions.

Inference: In the deployed state, the model receives new data and applies its learned patterns to make predictions. The inference process is facilitated through an API, which allows easy integration with other applications or services. The outcome of the inference process identifies the authenticity of the news content. This result can be used to flag or take action against fake news articles.

A. *Data Description*

FactWatch is a verified signatory to the International Fact-Checking Network (IFCN) and Meta's only Third-Party Fact-Checking Partner from Bangladesh . Thus, it is an important resource for collecting consequential and impactful Bengali fake news circulating in the Bangladeshi information ecosystem, particularly on Facebook. Facebook is the most popular social media in Bangladesh , and therefore the most important platform for the study of online fake news in Bangladesh. Since 2020, FactWatch has been a Third-Party Fact-Checker for Facebook, working independently to identify, analyze, and act on rumors and disinformation on the platform. They use AI resources from Facebook and their own methodologies for verification. Identified false content is then reduced in circulation by Facebook and tagged with warnings. FactWatch also produces explanatory articles and reports as part of its broader activities, independent of its agreement with Facebook, and has aligned its rating system with Facebook's for consistency.

The dataset, sourced from the FactWatch website, was automatically scraped and encompasses a comprehensive collection of 1,014 fact-checking reports, diligently covering the timeframe from October 4, 2021, to May 25, 2023. The curated dataset from Facebook focus on pages and groups notorious for spreading viral misinformation, as indicated by Facebook's recommendations. Selection criteria hinged on the fact-checkability using online sources, virality, and the potential for serious consequences like religious or political unrest.

In Table I, we present a comparison with other research works. For instance, authors in [26] concentrated on public opinion analysis on news through the lens of YouTube comments from 2017 to 2023, with the objective of enhancing sentiment analysis and facilitating the development of Bengali language models. Another significant study [27] explored the engagement with religious misinformation on Facebook, meticulously examining specific incidents of violence in Bangladesh that highlighted the profound impact of social media in propagating religious misinformation and its tangible consequences on societal harmony. In contrast, our dataset broadens the scope by encapsulating a wide array of misinformation themes across social media platforms, as evidenced by our dataset collated from FactWatch spanning from October 2021 to May 2023. Our work is distinguished by its comprehensive nature, encompassing meticulously annotated data that is not only pivotal for NLP and automated fact-checking efforts but also instrumental for in-depth misinformation studies. The dataset's extensive annotations are designed to facilitate nuanced analyses and model training, particularly addressing the challenges posed by the Bangla language and the specific cultural context of Bangladesh.

Table II outlines the column names and their descriptions for the Fact-Checking Report Dataset. This table outlines the key elements of each report in the dataset, including Post_ID, Post_Title, Post_Content_Summary, Original_Content_Text, Post_Rating, Investigation_Description, and Category, providing a clear understanding of the data's structure and content. More specifically, each report has a unique Post_ID, a title (Post_Title), and a summary (Post_Content_Summary) as presented on FactWatch's website. It also includes the Original_Content_Text, which are the fact-checked fake statements from various sources. The Post_Rating denotes the fact-checker's assessment of the news.

FactWatch uses several ratings for fact-checking: "False" for baseless information or content that is misrepresented as proof for unrelated events; "Altered" for misleadingly edited or AI-created content; "Partly False" for minor factual errors in otherwise accurate information; "Missing Context" for content that could be misleading without additional evidence; "Satire" for humorous or exaggerated content often misunderstood as factual; "True" for verified, accurate content; and "Unverified" for claims that remain inconclusive after fact-checking. These ratings help

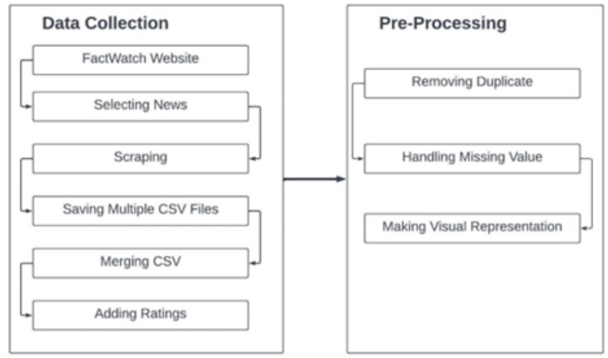

Figure 2. Flowchart Illustrating the Data Collection and Pre-Processing.

categorize the nature and integrity of various information types. The Investigation_Description provides a detailed account of the fact-checking process. Lastly, each report is categorized (Category) based on the theme of the fake news story.

B. Data Collection and Preprocessing

The data collection process in our research, detailed in Fig. 2, began with selecting news from the FactWatch website. The second step involved scraping the chosen news items, followed by saving the extracted data into multiple CSV files.

In the process of data collection, we employed scraping techniques using libraries such as Requests and Beautiful Soup to navigate and extract information from the FactWatch website. Additionally, Selenium was utilized to handle dynamic content and interactive elements on the site. To ensure the robustness of our data collection process, we implemented diligent error handling and periodically updated CSS selectors to adapt to any potential changes in the website's structure. To mitigate sampling bias, we employed strategies such as stratified sampling with the assistance of pandas, ensuring proportional representation across various report categories. Random sampling methods were also employed to secure a representative dataset. To address selection bias, we clearly defined and implemented inclusion and exclusion criteria within our data collection scripts, utilizing pandas for precise data filtering. Temporal bias was managed through rigorous monitoring of website changes to ensure that our data accurately reflected temporal trends in misinformation. After collecting the data, these CSV files were merged into a single file. The final part of the process included categorizing the data into various classes. This methodology, from news selection to data categorization, was instrumental in maintaining data integrity and is thoroughly represented in the flowchart in Fig. 1, showcasing our structured approach to data collection and organization. Below is a detailed overview of the organized process:

TABLE I. Comparison with other research works.

Paper	Focus	Data Source	Period Covered	Main Contribution
[26]	Public opinion analysis on news	YouTube comments	2017-2023	Dataset for sentiment analysis and language model development in Bengali
[27]	Engagement with religious misinformation	Facebook comments	Specific Usecase: the Nasirnagar violence in 2016, the Rangpur violence in 2017, the Narail protest in 2019, the Comilla violence in 2020, and the Sunamganj violence in 2021	Insights into social media's role in spreading religious misinformation
Our Dataset [28]	Misinformation on social media	FactWatch, social media content	October 2021 - May 2023	Broad Spectrum of Misinformation Types and Rich Annotations and Metadata. Comprehensive dataset annotated with fact ratings for combating misinformation, offering a rich resource for NLP, automated fact-checking, and misinformation studies, with extensive annotations aiding in deeper analyses and model training.

TABLE II. Column name and its Descriptions of Fact-Checking Report Dataset.

Column Name	Description
Post_ID	This refers to the designated unique identification number attributed to each of the fact-checking articles in the corpus.
Post_Title	This refers to the title of the fact-checking article.
Post_Content_Summary	This refers to the summary of the fact-checking article as given in FactWatch's website.
Original_Content_Text	This refers to the fake statements as found on social media or other internet sources (news, blogs, videos etc), which were fact-checked.
Post_Rating	This refers to the fact-checker designated rating (i.e False, Missing Context, Altered etc.) of fake news.
Investigation_Description	This refers to the descriptive text of the fact-checker's investigative process of debunking the fake news content.
Category	This refers to the classification of themes of fake news stories.

- **Selection of News Source:** The first step involved choosing the FactWatch website, selected for its comprehensive and reliable news coverage. This platform is crucial for accessing a wide range of news stories, ensuring our dataset includes diverse and significant news items.
- **News Item Selection:** Due to the vast amount of news on FactWatch, we focused on the most relevant news stories. This selective approach aimed to maximize the representativeness and relevance of our dataset.
- **Data Scraping Process:** We employed Python as our primary tool for data scraping, leveraging the Beautiful Soup library with the Chromium browser to navigate and meticulously extract information from each chosen news item. The data, including key elements such as news content, publication date, and

categorical classification, was systematically saved into multiple CSV files. Subsequently, these files were consolidated into a single comprehensive file, facilitating a more efficient and coherent analysis process.

- **Adding Ratings and Finalizing the Dataset:** The final step was categorizing each news item in the dataset into relevant classes of ratings. This categorization was crucial to organize the data effectively for subsequent analysis.

C. *Model Training and Validation Process*

We employ two deep learning models and complex linguistic patterns: LSTM and BERT with the Bangla base model.

Long Short-Term Memory (LSTM) [1] is an advanced RNN approach that solves the vanishing gradient issue found in conventional RNNs. Its structure allows it to capture long-range dependencies and subtle contextual details within text data, making it particularly suitable for pattern recognition in the sequential nature of news articles. This capability significantly enhances its performance in differentiating between genuine and fake stories. LSTMs are equipped with memory cells that retain and transport relevant information across long sequences, an essential feature for complex tasks in natural language processing.

Bangla BERT [2] is based on a Bengali language corpus for the Bengali language, embodying the advanced capabilities of the original BERT model. This model captures the bidirectional context of words by evaluating the text before and after a given word, leading to a more nuanced understanding of language intricacies. It is pre-trained on a large Bengali text corpus, including the Bengali Common Crawl and Wikipedia Dump Dataset, using the masked language model technique where it predicts intentionally hidden words to grasp deeper language context. The training uses the BNLP toolkit to develop a Bengali sentence piece model with a vocabulary size of 102,025. This architecture mirrors the base model, featuring 12 layers, 768 hidden units, 12 attention heads, and 110 million parameters, honed over 1 million steps with a Google Cloud GPU. In benchmarks like sentiment analysis, hate speech detection, and news categorization, Bangla BERT Base sets new standards, surpassing the results of multilingual BERT and Bengali Electra models.

Our classification workflow commences with the collection of data from a fact-checking website, followed by a meticulous data preprocessing process. The textual Bengali corpus news information is then encoded using the Bangla BERT model to prepare it for the training phase. For both LSTM and BERT, we meticulously configure hyperparameters—including class mode, validation splits, pooling methods, activation functions, optimizers, loss functions, as well as the number of epochs and batch sizes—to tailor the models for the specific task of misinformation classification.

D. *Performance Evaluation*

Following the completion of our dataset training using the LSTM and Bangla BERT models, we undertook a comparative analysis to gauge their performance. In the experimental analysis, we use precision, recall, and accuracy metrics. Following are the equations of each metric where, TP represents true positive, TN as true negative, FP as false positive, and FN as false negative.

The precision formula is given by:

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

The recall formula is defined as:

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

The accuracy formula is as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

4. Results and Discussion

The training of both LSTM and BERT models leveraged the capabilities of Google Colab, a cloud service that provides a robust and adaptable platform for conducting machine learning and deep learning research. Google Colab grants free access to formidable computational resources, including GPUs and TPUs, facilitating an accelerated training process for data-intensive models like the ones we employed. To evaluate the model's consistency and reliability, we employed a 5-fold cross-validation technique. This approach allowed us to iteratively train and validate the model across different subsets of the data, ensuring a comprehensive assessment. Additionally, we meticulously fine-tuned critical hyperparameters including the number of epochs, learning rate, and dropout rate to achieve the best possible model performance. Table III delineates the detailed hyperparameters applied in training the LSTM and BERT models, which are vital for steering the learning process and significantly influence the models' precision in news classification.

In the training of our LSTM and BERT models, we configured the class mode to 'Categorical' for our multi-class classification tasks. We allocated 10% of our dataset for validation purposes, ensuring that the model could be evaluated against unseen data during training. Max-pooling was utilized as our pooling technique, effectively reducing input dimensionality and capturing the essence of input features within binned sub-regions. We employed a sigmoid activation function, which is ideal for binary classifications, outputting probabilities between 0 and 1. The Adam optimizer was selected for its adeptness at managing sparse gradients and noisy data, which is typical of text datasets. We used binary-cross-entropy as our loss function, suitable for binary outcomes where the model

TABLE III. Model Hyperparameters

Hyperparameters	LSTM	BERT
Class Mode	Categorical	Categorical
Training Split	80%	80%
Test Split	20%	20%
Pooling	Max-pooling	Max-pooling
Activation Function	Sigmoid	Sigmoid
Optimizer	Adam	Adam
Loss Function	Binary-cross-entropy	Binary-cross-entropy
Epochs	100	100
Batch-size	15	15

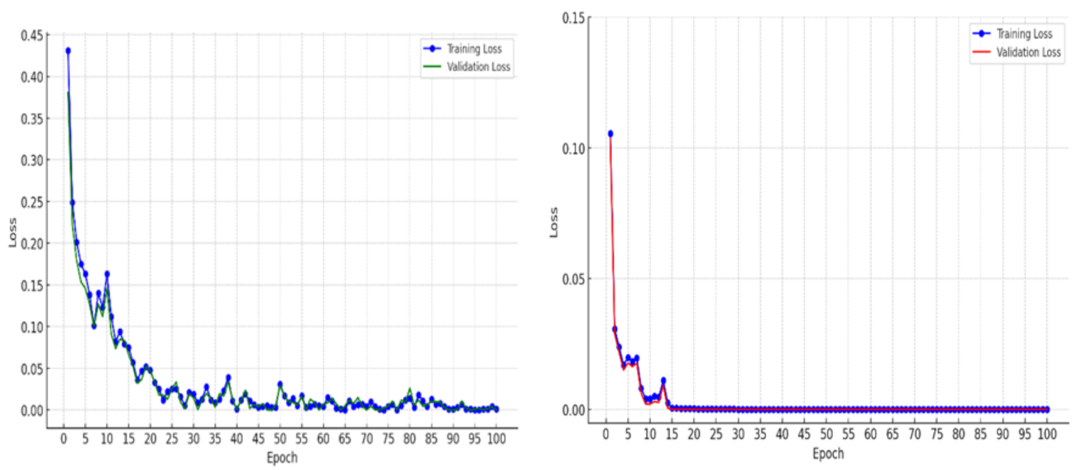

Figure 3. Loss curve of (a) LSTM and (b) BERT model.

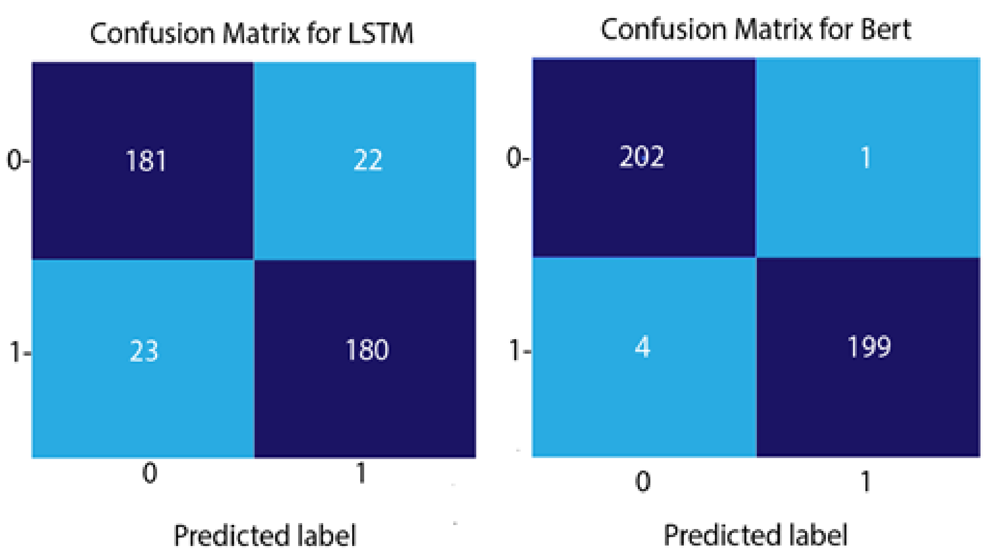

Figure 4. Confusion Matrix of (a) LSTM and (b) BERT model.

estimates the likelihood of an article being authentic or fake. Our models underwent training for 100 epochs, providing ample iterations for the models to discern complex patterns in the data, with a batch size of 15 to specify the number of examples processed in a single batch. Figure 3 illustrates the training and validation loss for LSTM and BERT model.

TABLE IV. Performance Measures of LSTM and BERT Model

Model	Precision	Recall	Accuracy
BERT	99.51%	98.06%	98.77%
LSTM	89.16%	88.73%	88.92%

To effectively evaluate the performance of our models, a confusion matrix was utilized, providing clear insights into their predictive accuracy. Figure 4 illustrates the Confusion Matrix of LSTM and BERT model. The BERT model's performance in the classification task was notably strong, as highlighted by the confusion matrix results. With 202 true positives and only 1 false positive, the model demonstrated a high level of precision, about 99.51%, showcasing its capability to correctly identify fake news with remarkable accuracy. The recall rate stood at a robust 98.06%, indicating the model's effectiveness in capturing the vast majority of actual fake news instances. The accuracy of the BERT model was also high at 98.77%, reflecting its consistent ability to make correct predictions across both classes.

On the other hand, the LSTM model, while still performing well, fell short of the BERT model's benchmark. It secured 181 true positives but had a higher rate of false positives at 22, leading to a precision of 89.16%. This suggests that the LSTM model was less accurate in its predictions of positive instances. Its recall rate was lower as well, at 88.73%, meaning it failed to identify all the positive instances that it should have. With an overall accuracy of 88.92%, the LSTM model was less adept at distinguishing between fake and real news compared to the BERT model, as evidenced by a higher number of both false positives and false negatives. These results underline the superior performance of the BERT model in this context, particularly in its reliability and consistency in fake news detection.

Considering experimental analysis, the BERT model outperforms the LSTM model in detecting fake news in Bengali. This could be attributed to BERT's transformer architecture, which can capture contextual dependencies in text more effectively than LSTM's sequential processing. The superior performance of BERT, with higher precision, recall, and accuracy, suggests that it is better at understanding the nuances of Bengali, likely due to its bidirectional nature and pre-training on a large corpus. The methodology of fine-tuning BERT, as opposed to training LSTM from scratch, may also contribute to the difference in performance. BERT benefits from transfer learning, where it adapts knowledge from a vast dataset to the specific task, providing a more robust starting point for classification tasks. These models guide us to develop an automatic news detection system in Bangla languages with complex structures like Bengali. The pre-trained models like BERT could be more effective for such tasks than traditional RNN-based models like LSTM. Future research could explore the integration of both models or investigate hybrid approaches that combine the sequential understanding of LSTM with the contextual awareness of BERT.

5. Conclusions and Future Work

Our research aimed to develop a deep learning-based approach capable of identifying online misinformation from news in Bengali on social media. Through trials and continuous refinement, we achieved a model with promising classification capabilities. BERT, with its contextual understanding, outperformed LSTM, leading us to choose it for our final system. While we faced dataset limitations, innovative solutions such as synthetic data generation were employed to enhance performance. In future work, we plan expand our dataset and refine the model to recognize the spectrum of news veracity, improving the tool's effectiveness and contribution to Bengali natural language processing.

The research has made significant strides in improving the detection of misinformation from news in Bengali from social media platform. Considering, limitation of the current approach, advance the model's classification ability to recognize varying degrees of information accuracy, such as partially true or biased content, which reflects the complex nature of real-world news more closely. Develop the model to identify multiple categories of news, moving beyond the binary true/false classification to include categories like satire, opinion, and analysis. Incorporate advanced language processing techniques to better grasp the intricacies and regional variations within the Bengali language, which can be pivotal in context interpretation

References

[1] B. Mateus, M. Mendes, J. Farinha, R. Assis, and A. Cardoso, "Comparing lstm and gru models to predict the condition of a pulp paper press," *Energies*, vol. 14, no. 21, p. 6958, 2021. [Online]. Available: https://arxiv.org/abs/1511.04108

[2] 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," *IEEE Access*, vol. 10, pp. 91 855–91 870, 2022.

[3] K. Sharifani, M. Amini, Y. Akbari, and J. Aghajanzadeh Godarzi, "Operating machine learning across natural language processing techniques for improvement of fabricated news model," *International Journal of Science and Information System Research*, vol. 12, no. 9, pp. 20 – 44, 2022. [Online]. Available: https://ssrn.com/abstract=4251017

[4] M. F. Mridha, A. J. Keya, M. A. Hamid, M. M. Monowar, and M. S. Rahman, "A comprehensive review on fake news detection with deep learning," *IEEE Access*, vol. 9, pp. 156 151–156 170, 2021.

[5] A. Thota, P. Tilak, S. Ahluwalia, and N. Lohia, "Fake news detection: a deep learning approach," *SMU Data Science Review*, vol. 1, no. 3, p. 10, 2018. [Online]. Available: https://scholar.smu.edu/datasciencereview/vol1/iss3/10/

[6] N. Aslam, I. U. Khan, F. S. Alotaibi, L. A. Aldaej, and A. K. Aldubaikil, “Fake detect: A deep learning ensemble model for fake news detection,” *Complexity*, pp. 1–8, 2021.

[7] 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)*, 2021, pp. 1–8.

[8] P. Meesad, “Thai fake news detection based on information retrieval, natural language processing and machine learning,” *SN Computer Science*, vol. 2, p. 425, 2021.

[9] J. A. Nasir, O. S. Khan, and I. Varlamis, “Fake news detection: A hybrid cnn-rnn based deep learning approach,” *International Journal of Information Management Data Insights*, vol. 1, no. 1, p. 100007, 2021.

[10] R. K. Kaliyar, A. Goswami, and P. Narang, “Fakebert: Fake news detection in social media with a bert-based deep learning approach,” *Multimedia Tools and Applications*, vol. 80, no. 8, p. 11765–11788, 2021.

[11] S. B. S. Mugdha, S. M. Ferdous, and A. Fahmin, “Evaluating machine learning algorithms for bengali fake news detection,” in *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, 2020.

[12] V. V. Hirlekar and A. Kumar, “Natural language processing based online fake news detection challenges – a detailed review,” in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, 2020, pp. 748–754.

[13] D. S and B. Chitturi, “Deep neural approach to fake-news identification,” *Procedia Computer Science*, vol. 167, pp. 2236–2243, 2020.

[14] M. G. Hussain, M. R. Hasan, M. Rahman, J. Protim, and S. A. Hasan, “Detection of bangla fake news using mnb and svm classifier,” in *2020 International Conference on Computing, Electronics Communications Engineering (iCCECE)*, 2020, pp. 81–85.

[15] M. Hossain, M. Rahman, M. Islam, and S. Kar, “Banfakenews: A dataset for detecting fake news in bangla,” *arXiv preprint*, 2020. [Online]. Available: https://arxiv.org/abs/2004.08789

[16] V. M. Krešňáková, M. Sarnovský, and P. Butka, “Deep learning methods for fake news detection,” in *2019 IEEE 19th International Symposium on Computational Intelligence and Informatics and 7th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics (CINTI-MACRo)*, 2019, pp. 000 143–000 148.

[17] P. Bharadwaj and Z. Shao, “Fake news detection with semantic features and text mining,” *International Journal on Natural Language Computing (IJNLC)*, vol. 8, no. 3, 2019. [Online]. Available: https://ssrn.com/abstract=3425828

[18] M.-Y. Chen, Y.-W. Lai, and J.-W. Lian, “Using deep learning models to detect fake news about covid-19,” *ACM Transactions on Internet Technology*, vol. 23, no. 2, p. Article 25, 2023.

[19] H. Najadat, M. Tawalbeh, and R. Awawdeh, “Fake news detection for arabic headlines-articles news data using deep learning,” *International Journal of Electrical Computer Engineering (2088-8708)*, vol. 12, no. 4, 2022.

[20] A. Ali and A. Abdullah, “Fake accounts detection on social media using stack ensemble system,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 3, pp. 3013–3022, 2022.

[21] F. Benabbou, H. Boukhouima, and N. Sael, “Fake accounts detection system based on bidirectional gated recurrent unit neural network,” *International Journal of Electrical Computer Engineering (2088-8708)*, vol. 12, no. 3, 2022.

[22] P. Mookdarsanit and L. Mookdarsanit, “The covid-19 fake news detection in thai social texts,” *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 2, pp. 988–998, 2021.

[23] S. Khaled, N. El-Tazi, and H. M. O. Mokhtar, “Detecting fake accounts on social media,” in *2018 IEEE International Conference on Big Data (Big Data)*, 2018, pp. 3672–3681.

[24] K. R. Purba, D. Asirvatham, and R. K. Murugesan, “Classification of instagram fake users using supervised machine learning algorithms,” *International Journal of Electrical and Computer Engineering*, vol. 10, no. 3, p. 2763, 2020.

[25] Y. Boshmaf, D. Logothetis, G. Siganos, J. Lería, J. Lorenzo, and M. e. a. Ripeanu, “Íntegro: Leveraging victim prediction for robust fake account detection in large scale osns,” *Computers Security*, vol. 61, pp. 142–168, 2016.

[26] L. H. Chowdhury, S. Islam, and S. Shatabda, “A bengali news and public opinion dataset from youtube,” *Data in Brief*, p. 109938, 2023.

[27] M. S. Al-Zaman and M. M. S. Noman, “A dataset on social media users’ engagement with religious misinformation,” *Data in Brief*, vol. 49, p. 109439, 2023.

[28] M. R. A. Rashid, R. Roy, N. Monsoor, M. A. Saleh, A. A. H. Khan, M. A. Rayhan, K. Foysal, and D. M. S. Rahman, “Fact-checked facebook news corpus from bangladesh,” 2023.