Fake News Detection: Performance Analysis of RNNs and Transformer Approaches

By
Suchana Akter
211-15-14623

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the **Degree of Bachelor of Science in Computer Science and Engineering**

Supervised by

Dr Naznin Sultana
Associate Professor
Department of Computer Science and Engineering Daffodil International University

Co-Supervised by

Jamilul Huq Jami
Lecturer
Department of Computer Science and Engineering Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

Sep 17, 2025

APPROVAL

This Project titled **"Fake News Detection: Performance Analysis of RNNs and Transformer Approaches,"**, submitted by **Suchana Akter,** ID No: **211-15-14623** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **17 September, 2025.**

BOARD OF EXAMINERS

Dr. S.M Aminul Haque — **Chairman**
Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Ms. Nazmun Nessa Moon — **Internal Examiner**
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Abbas Ali Khan — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Zulfiker Mahmud — **External Examiner**
Professor
Department of Computer Science and Engineering
Jagannath University

DECLARATION

We hereby declare that this project has been done by us under the supervision of Dr **Naznin Sultana**, **Associate Professor**, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this projectnor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Naznin Sultana
Associate Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by:

Jamilul Huq Jami
Lecturer
Department of Computer Science and Engineering
Daffodil International University

Submitted by:

Suchana Akter 17/09/25

Suchana Akter
Student ID: 211-15-14623
Department of Computer Science and Engineering
Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project (FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Naznin Sultana**, **Associate Professor**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **NLP** to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting themat all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Cultural & linguistic challenges The rise of fake news spreading like wildfire across digital platforms, particularly at the low-resource level in languages such as Bangla, where fact-checking tools are scarce, has become a matter of grave concern. To solve this problem, we are investigating content-based automated detection systems developed with deep learning and transformer approaches.

In this research, we test four types of deep learning models (LSTM, GRU, BiLSTM, BiGRU) and one transformer model (mBERT), using a Bangla fake news dataset, which was gathered from several news sites, journalism platforms, and social media. The data was pre-processed by standardization, tokenization, and balancing and models were run for measuring accuracy, precision, recall, and F1-score.

Results indicate that the best accuracy (97%) was obtained with mBERT, which was superior to all RNN-based models. Among the RNNs, BiGRU fare the best with 95%, followed by LSTM, GRU, and BiLSTM, each at close to 94%. This horizontal contrast confirmed the best contextual comprehension derived from mBERT and the lowest misclassification rate.

These observations highlight the effectiveness of transformer based models such as mBERT in detecting Bangla fake news. In future, we will investigate domain-specific training, hybrid models, multilingual transfer learning and integration with dialects, code-mixed text and multimodal features to improve the scalability and real-world use-case coverage.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

But then came the digital age and the emergence of misinformation which is also downright an open global challenge dominating over public opinion, on social stability, political scenario and seems to be fuelling itself through social media, digital news portals, user generated content and so on, spreading misinformation much more fast than it has ever crossed boundaries of being able to differentiate valid news from the faulty one. The rapid consumption and spread of unverified news, then lead to the provision of disinformation, manipulation in politics, turmoil in society. Efforts to automate this detection are therefore challenged by the fact that standard fake news detection apparatuses are severely under-represented for high resource languages such as English in the fake news and misinformation literature.

Bengali, one of the top 10-most spoken languages worldwide, is spoken by over 270 million people primarily in Bangladesh and some parts of India. Social media and online news outlets within these domains have over the years propagated misinformation that has greatly influenced public opinions, political discourse and societal traditions. But fact-checking tools are scarce for the Bangla language, and manually confirming articles takes time. Therefore, a Fake News Retarding system should be implemented in an automated way to ascertain the truth of posted news in a more dynamic and accurate manner.

Previous work on fake news detection, rulebased methods, and statistical models have been proven to be insufficient with the different nuances in language, contextual drifts, and deceptive styles of writing. These approaches primarily rely on keyword matching and handcrafted features which may not be effective in detecting more complex misinformation strategies. Machine learning and deep learning, however, have shown their effectiveness with Natural Language Processing applications such as sentiment analysis, text classification, and of course, identifying/fighting fake news. Some deep learning approaches such as Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM) and Gated Recurrent Units (GRU), have achieved very good results on fake news detection, but the ability of Recurrent Neural Networks (RNNs) to learn long-short distance dependencies is poor, and therefore, they would not be ideal for understanding more complex language structure and contextual relationships.

The transformer model, and its descendant BERT, has managed to rejuvenate fake news detection in terms of accuracy and generalization. Transformers employ selfattention mechanism to capture contextual relationships across full sentences.

Multilingual BERT (mBERT) is a pretrained transformer model for a variety of languages including Bangla which is useful for misinformation detection task in low-resourced languages. With context-aware learning and bidirectional encodings, mBERT are 1 stage over than the old RNN-based models in link detection accuracy.

1.2 Problem Statement

Such detection of fake news in bangLa has received very little serious research attention, and automatic and reliable systems are conspicuously missing in the wake of the proliferation of misinformation in bangLa. While conventional machine learning approaches entail the construction of customized features from the language, however, these do not encompass the full deviations of the deceptive-text writing style. For deep learning models like LSTM, GRU, BILSTM, BIGRU performance is quite good, but there are still some weak points including long-term dependency and contextual understanding. In addition to this, there are several other challenges related to the imbalance in the datasets, code-mixed texts (Bangla-English), or the unavailability of data for labelling, which further hampers the detection of fake news in practice.

This research aims at developing a modern fake news detection system approach with deep learning and transformer-based models, where it will primarily focus at comparing LSTM, GRU, BiLSTM, BiGRU, and mBERT. This way, the project will serve as a future research, which one of these models can provide the maximum accuracy and generalization to fake news detection task in Bangla. This work will be of great interest by being insightful into the capabilities (and limitations) for both RNN-based and transformer-based models for future research on NLP-based misinformation detection for low-resource languages.

1.3 Research Objectives

Addressing the problem of fake news detection in Bangla, the early works explore deep learning and transformer based models. The prominent contribution of the novelty project is to compile a large scale Bangla fake news dataset from trustworthy news agencies, independent journalism websites, and some other social media sources by collecting and pre-processing the data. Some data-cleansing operations like tokenization, padding, normalization and, stop word removal are performed in order to enhance the quality of the datasets; and oversampling and under sampling work is done to reduce the effects of the class-imbalanced problem. In the process of model development and training, deep learning models were developed and trained, so that LSTM, GRU, BiLSTM, BiGRU model, as well as transformer-based mBERT model with optimized hyperparameters, batch size, learning rates, training epochs were produced and used for the high efficiency.

The performance was confirmed by performing the comparison between these models with important criteria such as accur acy, precision, recall, F1-score, and confusion matrices; and revealing the misclassification patterns which of analysis limitations of

the models. Then, a comparison is made between deep learning based models and transformer based architectures and identifies which model best generalizes the task of Bangla fake news detection. For future improvement we refer to domain-specific pre-training, hybrid model architecture, multilingual transfer learning, and adversarial training as four possible techniques to explain higher accuracy. Each enhancement to detection capability will be coupled with an addition to the dataset from the multimodal misinformation sources in the form of text, images and videos.

1.4 Contributions of the Study

This research contributes significantly to Bangla fake news detection1. It provides for the first time large scale comparison of deep-learning models for LSTM, GRU, BiLSTM and BiGRU versus transformer based model (mBERT) for Bangla fake news classification. The balanced dataset is a high-quality one and made as a part the newly proposed balanced Bangla misinformation detection for effective model shaping and evaluation. Several preprocessing steps were performed, such as addressing class imbalance, tokenization and padding, and text normalization, to improve model effectiveness. The models were extensively tested for their performance with accuracy, precision, recall, F1-score and with confusion matrices. It also lays down the overall perils and boons of these RNN based models as compared to transformers and provides very crucial inputs for potential future works which could be done on fake news detection systems. Finally, the research pointed out the direction of future work on hybrid deep learning architectures, multilingual transfer learning and multimodal fake news detection so that more accurate and efficient misinformation detection solutions can be developed.

1.5 Organization of the Report

The structure of the thesis is organized as follows:

A Chapter 2 provides a detailed literature review on fake news detection techniques including traditional machine learning, deep learning, and transformer-based techniques.

B. Chapter 3 provides detail about the research methodology, such as data collection, pre-processing methods, model architectures, experimental setting and evaluation metrics.

C.- Results and Analysis The results are discussed in Chapter 4 including a comparative performance analysis among the LSTM, GRU, BiLSTM, BiGRU and mBERT model.

D. Finally Chapter 5 concludes the work and presents the future scope of research work, listing its findings and discussing the future enhancement in the field of Bangla fake news detection.

With this comprehensive examination, our effort is to fill this research void in the Bangla fake news detection space and present a general NLP solution for such detection that is both effectual and scalable and can be applied on other languages.

Chapter 2

Background

2.1 Introduction

Majority of the academia in the domain of fake news detection for Bangla was focused on the use of traditional ML model such as SVM, Naïve Bayes or Decision Trees and could produce only moderately effective result. The recently emerged deep learning techniques improved the predictive performance of some models such as LSTM, BiLSTM, GRU, and CNN significantly, while hybrid architectures, including CNN-LSTM, outperform them. Transformer-based models like Bangla-BERT and mBERT were proposed as the most successful due to surpassing the recurrent neural networks in context understanding.

A lot of the researchers worked on constructing massive Bangla fake news datasets along with some pre-processing tasks such as tokenization, normalization, and class balancing. Those would then be judged in terms of metrics of interest such as accuracy, precision, recall, and F 1 score, which consistently displayed the BERT models as the top performers. However, there are several issues unresolved, in- cluding how to address constrained data sets, class im- balance, and insufficient use of fine-tuned transformer models.

The majority of prior researches are based either on headline classification [30, 18]; or not on the full text [6]; or not on the realtime detection [20]. This study seeks to address such gaps by tuning mBERT and BiGRU, handling data imbalance, and conducting a thorough comparison with models in order to enhance the effectiveness of Bangla fake news detection.

2.2 Literature Review

Mondal et al. studied the effect of fake news in Bangla-speaking region and proposed a deep GRU model, as a viable solution for fake news detection. The study applied a number of pre-processing steps including lemmatizing and tokenization, as well as balancing class distribution through oversampling, which resulted in containing 58,478 news items. This is evidenced by the excellent 94% accuracy on classifying Bangla fake news obtained by the GATED-based model over many of its current competitors. The work provided a detailed discussion about data preparation, model selection, training, and evaluation, and used precision, recall, \emph{F1 score}, and accuracy as evaluation metrics. The significant contributions include not only a large Bangla fake news dataset and a high performing deep model, but also insight into how GRU shines at its skill of squashing misinformation.

In this Work [2] authors have created a Bangla Fake News corpus, compiling data from existing sources and new data collection, resulting on a pre-compiled accessed set of 4678 distinct news articles. They make a comparison with various machine learning methods such as Logistic Regression, Support Vector Machine, K-Nearest Neighbours, Multinomial Naïve Bayes, AdaBoost, and Decision Tree. They have also played with deep learning models namely Long Short-Term Memory, Bidirectional LSTM, Convolutional Neural Networks, Hybrid CNN-LSTM, CNN-BiLSTM and transformer-based models Bangla-BERT, mbert (multilingual BERT). Of the models though CNN got highest accuracy of 95.9%, followed by CNN-LSTM with 95.5% and BiLSTM with 95.3%, thus attesting such efficiency of the deep learning and transformer-based methods in classifying Bangla fake news.

In this Work [3] Hossain et al. explored the task of verifying Bangla fake news by training it on a dataset of 57,000 Bangla news articles annotated for the trustworthiness and the counterfeits. They did K-fold cross-validation before adopting BiLSTM with GloVe and FastText embeddings and attained 95% and 94% accuracy, respectively. Other experiments on GRU model obtained accuracy of 77% and the BiLSTM model at 96%, showing its strength in this sense.

In this paper [4], detection of bangla fake news has been carried out using supervised machine learning algorithms, and a world-wide concern is initiated for false news spreading using online and social media platforms. The paper considers Multinomial Naïve Bayes, Support Vector Machine classifiers with Count Vectorizer and TF-IDF Vectorizer for feature extraction. Under this approach, fake news was labeled based on the polarity of the article associated with it. Experimental data revealed that SVM with linear kernel performed better than the Multinomial Naïve Bayes reaching a 96.64% of accuracy. The work demonstrates that machinelearning based methods will be very beneficial for the Bangla fake new's identification and the research in this less-explored area should be pursued further.

In this review paper [5], the authors have examined how fake news spreads on the internet, with specific focus on the impact of misinformation, including yellow journalism, on the society and politics. They proposed to use AI-based models to detect fake news using machine learning and deep learning algorithms. The writers used and compared Individual Recurrent Neural Network, LSTM, BiLSTM, GRU versus a transformer-based model, i.e., BERT for fake news classification. BERT achieved the highest accuracy of 95%, followed by RNN (94.58%), BiLSTM (94.29%), GRU (93.22%) and LSTM (92.84%). The results demonstrate that BERT is better to learn the context dependence in text than other models. Therefore, it was the optimum model in detecting fake news in the Bangla language.

A growth of fast food - type online news paper would make it very difficult to identify the rubbis from real news. The headlines are shamelessly written to grab public readership. The are rooted in as much confusion and warped conception there as the headline is misleading. To mitigate this, the work of [6] developed a fake news headline classifier in Bengali. This classification model made use of the Gaussian Naïve Bayes technique over a newly curated data and stood tall as it achieved a performance of 87% on the precision front. This outperformed other methods by addressing an important void in the fake news detection work in Bengali. Although there exist automated methods for fake news detection on English newspaper articles, the paucity of such work has caused Bengali news to be underdeveloped in this domain.

The work in this paper [7] has solved this problem in South Asia with recognition for Data mining algorithms. There are roughly 200 million Bengali speakers in the world, so this is a significant task. The final result obtained by the Random Forest Classifier had a 85% accuracy, suggesting the possibility of the detection of fake news in the Bengali language. In the future, we can optimize more on classification for the precision and develop the real-time web-based architecture to combat misinformation fast. Online news reading is increasingly harder to distinguish between news and fake.

In this work [8], we contribute to analysing this proliferating case of fake Bangla news, identifying the pattern of language use separating real articles from fake ones. Given such data as this dataset with 48,678 real & 1,299 fake Bangla news, we utilize a deep learning model in order to deal with class imbalance and ensembling by using random under-sampling.

The authors in [9] dealt with the problem of imbalanced data in Bangla fake news detection. This work accordingly suggested easing information imbalance methods to better assist the conduction of classification. With that notion in mind, some techniques, including Smote and stacked generalization, were used to maximise classification precision on the BanFakeNews dataset, which is critically lesson-awkward. With growing deception on the internet, pressuring the localization of fake news will be crucial for societal smarts. The research work contributes in terms of the orientation of Bangla language towards fake news detection systems by insights on solutions of imbalanced data, and classifier performance optimization methods.

The burgeoning of fake news calls for strong detection mechanisms, notably in under-resourced languages such as Bangla. This work in [10] addresses these issues by a collection of large dataset containing 50,000 news articles and applying several deep learning models such as GRU, LSTM, CNN and hybrid architectures to detect fake news. We used performance metrics such as recall, precision, F1-score, and accuracy to measure the efficacy of the models. Moreover, the work emphasized the significance of dataset balancing and continual tuning as an enabler for better Bangla fake news detection and for compensating for resource paucity in low-resource language research.

2.3 Gap Analysis

For now, Bangla fake news detection work respond on machine learning model suffers model inability to understand the context in an efficient manner along with deep learning model, irrespective of class imbalance and data set limitation. While BERT-based models have emerged as promising, only a handful of works have been done for its fine-tuning forms for Bangla yet. Most of the previous works are focusing on the headline classification rather than analyzing in the context of the full article or detecting in real time. This work will therefore fill in gaps by finetuning mBERT and BiGRU addressing data skew, and extensive model comparison for betterment of Bangla fake news detection.

Chapter 3

Research Methodology

3.1 Methodology

3.1.1 Overview

This section explains the methodology used in this study that is illustrated in Figure 2.1 for text classification with deep learning-based models (LSTM, GRU, BiLSTM, BiGRU) and transformer-based model (mBERT). The methodology is composed by data collection, data preprocessing, model architectures, experimental setup, model training, and evaluation. All the elements are significant for the efficiency and reliability of the classification.

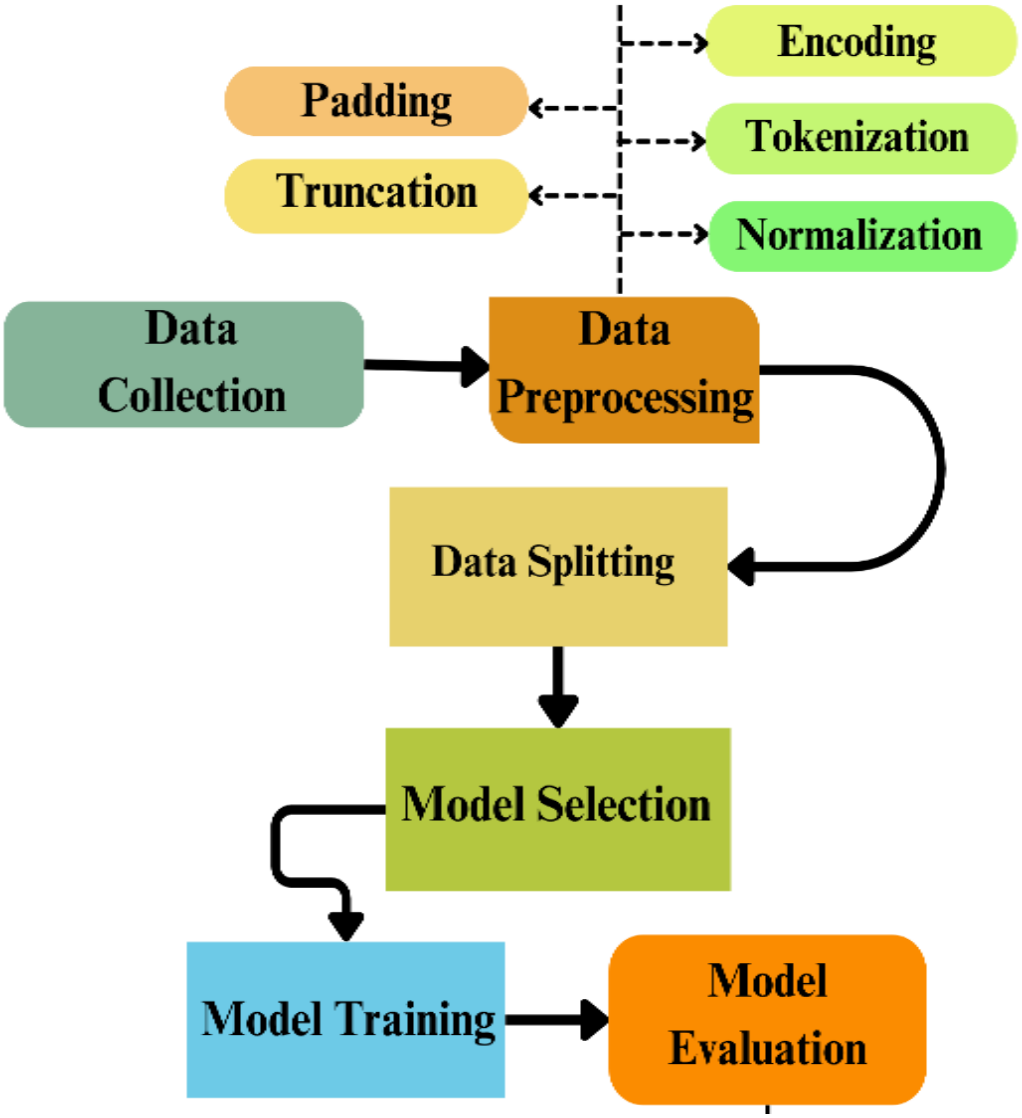

Figure 3.1: Proposed Methodology of Our Work

3.2 Data Collection

The dataset exploited in this project is represented in Figure 3.2 is crawled from different Bangla news sites and forum sites and from social networks. These included reliable news agencies, independent outlets and local sources and others. The whole dataset was divided into real and fake news.

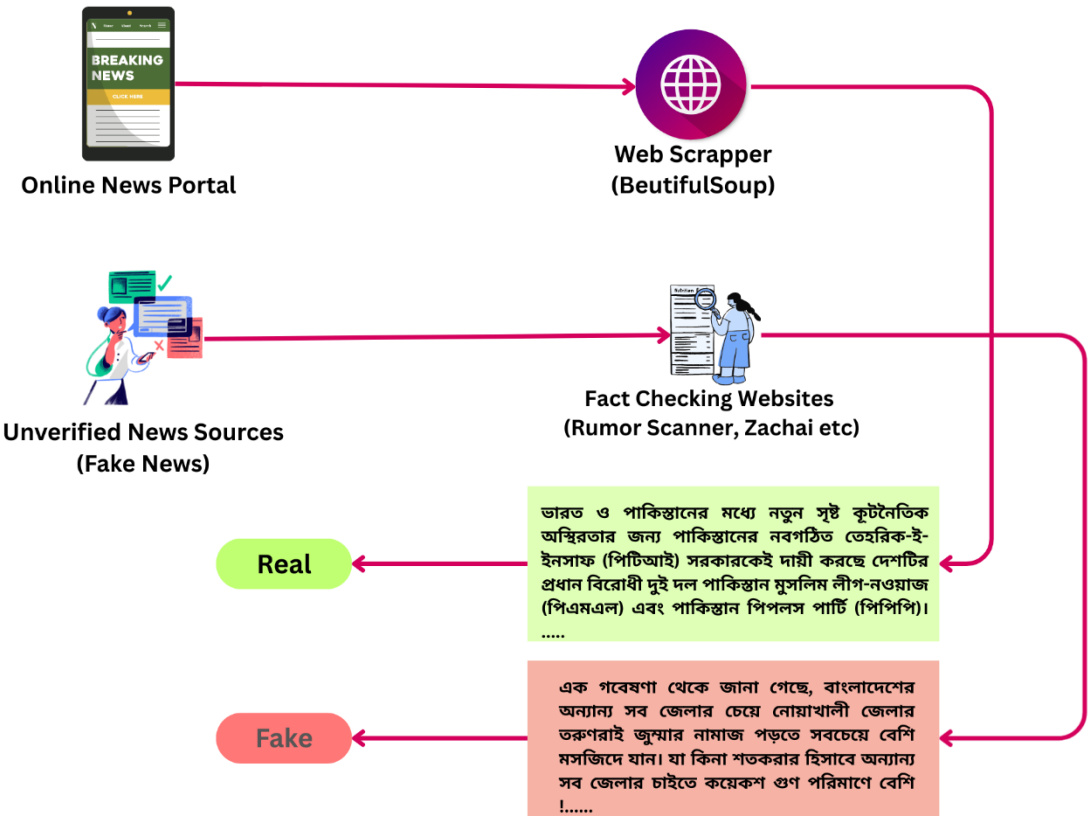

Figure 3.2: Data Collection Pipeline

	content	label
0	গত ৪ জুলাই আন্তর্জাতিক শান্তি সংস্থা ইউনেস্কো ...	fake
1	খেলা ডেস্ক আরটিএনএন আবুধাবি: এশিয়া কাপে নিজেদে...	real
2	বাসের সিট নিয়ে সত্য কথা বলায় হেল্পারের প্রেমে ...	fake
3	সূচক এবং লেনদেনের নিম্নমুখী প্রবণতার মধ্য দিয়ে...	real
4	শিল্পমন্ত্রী আমির হোসেন আমু যুব সমাজকে রক্ষায় ...	real
5	বাংলাদেশের টিভি পর্দায় আসছে আন্তর্জাতিক পুরস্ক...	real
6	খুবই নির্মম একটি ঘটনা। কেননা, শ্বশুরকে গাছে সঙ...	real

Figure 3.3: Sample of Our dataset

To verify fake news cases we referenced to several fact checking websites (e.g. Rumour Scanner, etc) to check the reliability of the classification. These fact-

checking services examine news and compares with well-documented cases and statements to determine whether or not it is true. At the beginning, the collection had 2,598 fake news and 16,850 true news articles, 19,448 in total.

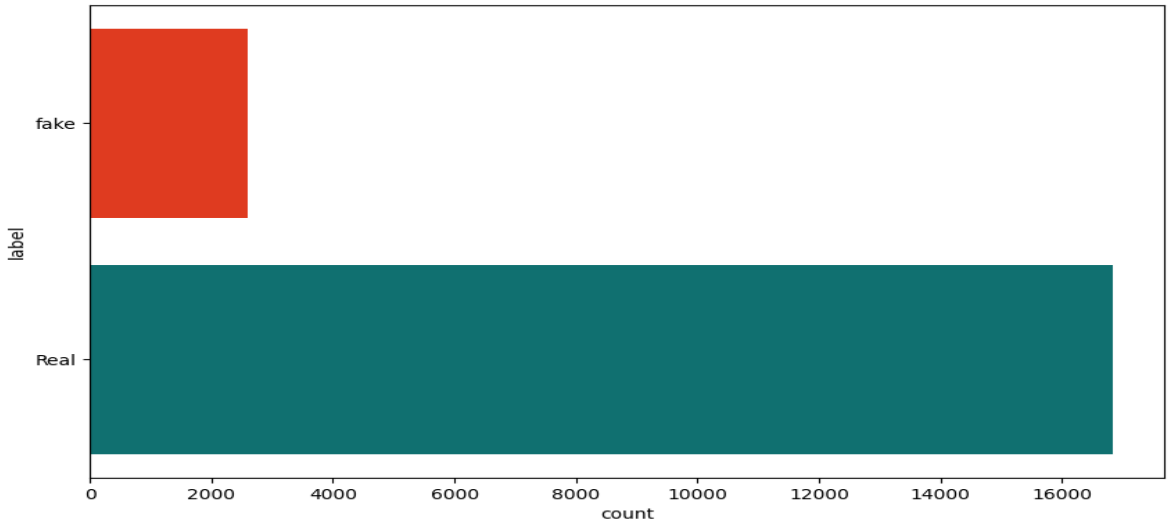

Figure 3.4: Data Distribution of Our Proposed Dataset

3.3 Data Pre-processing

Due to the class imbalance oversampling and undersampling methodologies were used in order to obtain a balanced dataset. RandomOverSampler to over sample the fake news samples to have the same number as the real news samples and the RandomUnderSampler to downsample the number of real news samples to a constant size. Hence, the both fake and real news categories were modified to be 3,000 samples each in Figure 3.4 (balanced by class (equal class distribution)).

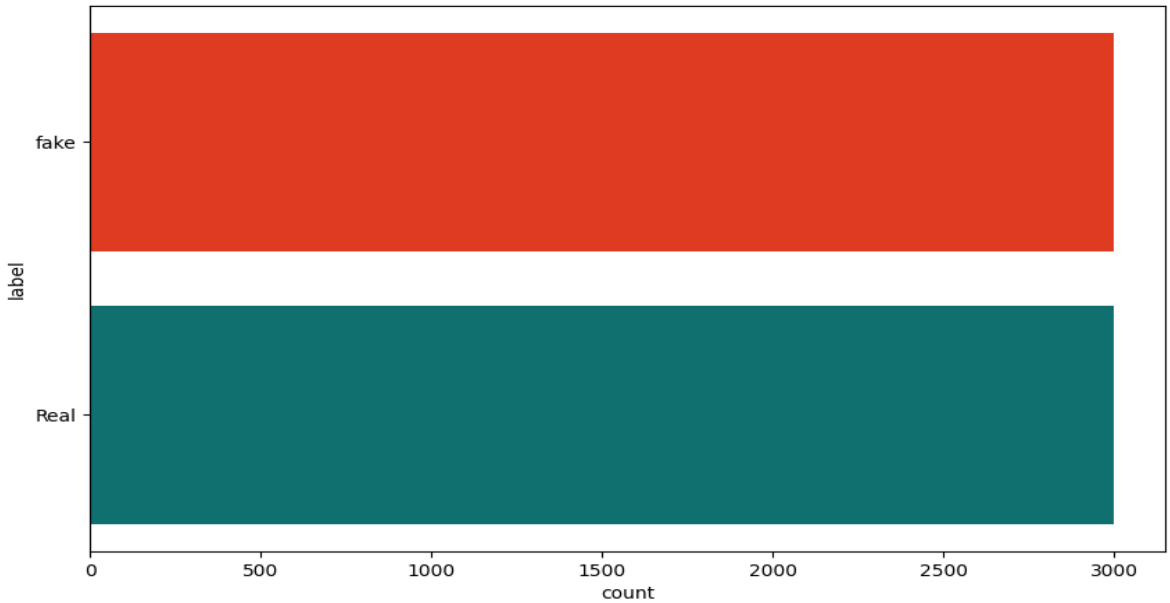

Figure 3.5: Data Distribution After Pre-processing

The balanced dataset reduced the risk of any biases towards the majority class such that classifiers were trained on an even number of real and fake news examples. The gathered text data was also pre-processed through a sequence of steps customized for deep learning models as well as transformer-based models. These prepossessing methods which aim to

improve feature representation and model performance.

Pre-processing for DL-based Models
Due to the fact that deep learning models only take in structured numerical input, preprocessing is a key step in transforming the text data into a form suitable for training. A number of pre-processing methods – described in Figure 3.5 were introduced to maintain consistency, reduce learning difficulty, and pick up useful feature representations. The following steps were employed:

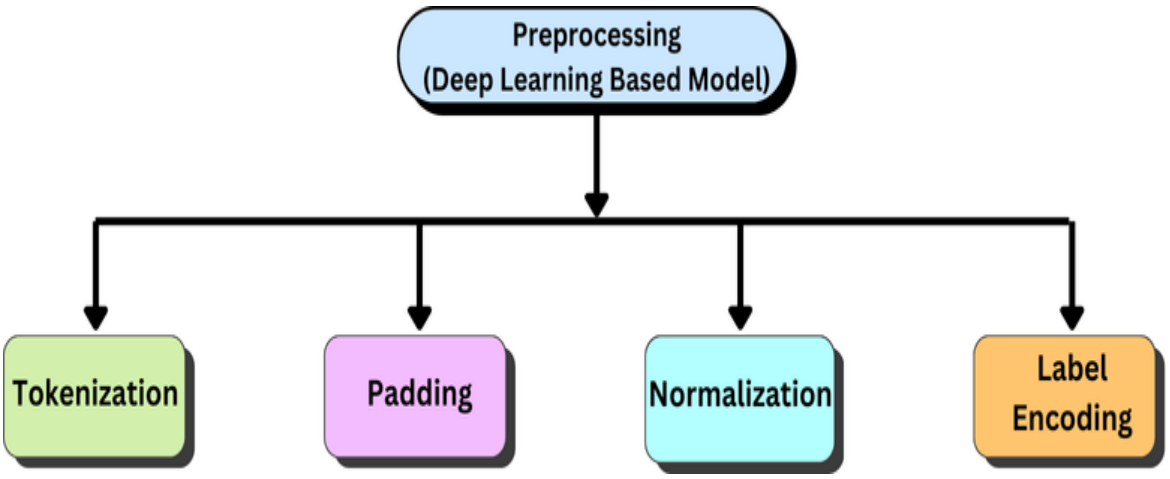

Figure 3.6: Pre-processing steps for Deep Learning Based Models

3.3.2.1 Tokenization

Text was tokenized at the subword level using the mBERT tokenizer. This method assists in dealing with OOV words by splitting them into more frequently observed parts and by this all (even previously unseen) words become meaningful to some extent. The tokenized sequences are mapped into integer IDs according to the pre-trained mBERT vocabulary.

3.3.2.2 Truncation & Padding

Since transformer models take in input of fixed length, the maximum length of any sequence was set to 128 in each case so as to have homogenity across all inputs. Even sentences that are much shorter were not longer than the maximum sequence length in order to not allocate more memory than needed, and shorter sentences were padded by the PAD token in order to have a uniform input shape which is important for running the data through the network in batches. The purpose was to maximize the compu tation feasibility and to preserve the significant textual information.

3.3.2.3 Normalization

Text Normalization Text normalization with the CSE-BUET NLP Normalizer was performed for better quality of the dataset and also to remove noise. All text were lowercased for consistency, and to prevent case sensitivity. Introduces abundant noise by dropping out unnecessary characters, punctuations and special symbols from the corpus, which is conductive to cleaner processing and model generalization. Stop words were here also discarded, as to retain only affecting words to the model learning. making the dataset very well structured and cleaner, reducing the extra, non-informative loads which would impair the performance of the model Instead of suspecting these words (choices are made as to how much it might take to clean other words starting with various prefixes), the dataset was made well structured and cleaner applying the CSEBUET NLP Normalizer.

3.3.2.4 Special Tokens

Transformers introduce special tokens to represent and separate contexts of different sequences. A special CLS token was prefixed to each sequence (as shown in Figure The unconditioned SEP was added at the end of each sequence to indicate disambiguation between textual inputs being separated into sentences (especially if the task is two-sentence or question-answering related). These tokens act as structural advices for the model to learn contextual relationships more easily.

3.3.2.5 Label Encoding

Class labels were converted into numerical labels for classification as per the Hugging Face framework. The class labels were in the same way as deep learning step. Labels were then restructured with this transformation, so that the model could process them, and they were usable with PyTorch-based training pipelines.

Through these preprocessing steps, the dataset was prepared for maximum efficiency computation in mBERT, which guaranteed a better feature representation and model efficiency. These mechanisms allowed the transformer-based model to learn fine-grained dependency for language patterns which contributed to the improved performance in text classification while being able to preserve more contextual information.

3.4 Model Architectures

In this work, we investigate various deep learning architectures, notably recurrent-based and Transformer-based models. Recurrent models are particularly suited for sequential data, capturing the dependencies in the input over time. On the contrary, transformer type architecture based models use self-attentions to better capture long range dependencies. The contrast between these architectures gives us more understanding about the strengths and drawbacks of them for the task.

3.4.1 Deep Learning Based Models

In this paper, four well-known recurrent-based deep learning models including LSTM, GRU, BiLSTM, and BiGRU were tested. These models are especially good at modelling sequential dependencies in textual data.

3.4.1.1 Long Short-Term Memory (LSTM)

LSTM is essentially a special type of RNN that is designed to overcome the problem of vanishing gradient incurred by RNN when investigating long-term dependencies in a sequence of inputs (Figure 3.7. LSTMs achieve this with memory cells and gates that regulate the flow of information. The main parts of an LSTM unit are a cell state or long-term memory and three main gates: a forget gate, an input gate, and an output gate. The forget gate receives both the previous hidden state and the current input and then it uses a sigmoid activation in order to decide what information should be kept and what should be thrown away. The input gate is basically deciding what new information we are going to store in the cell state. More concretely, in order to compute the amount of the memory cell to be updated at the current time step, an activation of sigmoid function is employed, while for computing candidate values that will be abandoned or stored in memory, an activation of tanh function is used. This time the output gate decides what to include in the cell state, deciding what of the cell state to let through as the next hidden state, both using a sigmoid, and tanh to squash the inputs. The LSTMs, operating in the realm of information flow, are particularly effective at capturing long-range dependency, and this

makes them particularly useful in NLP, speech recognition, and time series. LSTMs solve this by having gates to allow relevant short-term features to be stored and to forget irrelevant data, and in this way is able to find the right measure of short and long term dependence to learn. LSTMs also handle long-term dependencies much better than the vanishing gradient problem in standard RNNs.

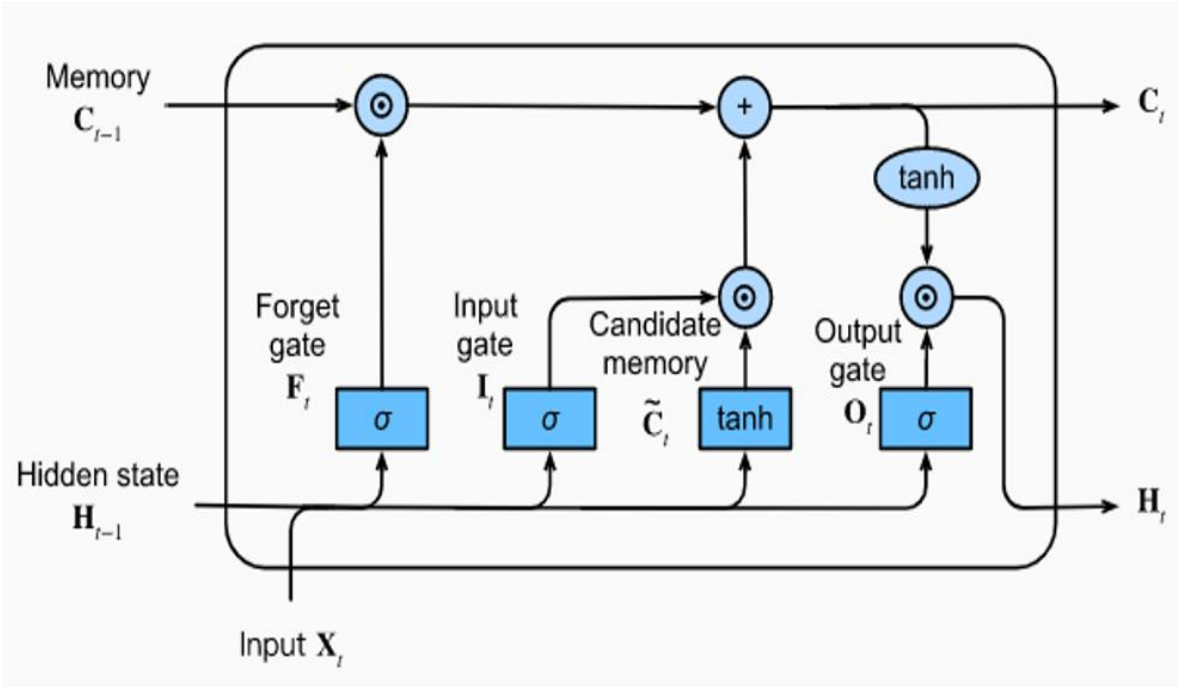

Figure 3.7: Architecture of LSTM Layer

Here this model is to perform binary classifications using the LSTM layers in the Sequential model in Figure 3.8. This architecture is also suitable for a variety of other tasks such as sentiment analysis, spam detection, and text categorization. The first unit in the sequence is the Embedding Layer, which would transform words to dense vector representation, taking max words (vocabulary) to a 100-dimensional embedding space and handling sequences of fixed length max sequence length and allow the model to learn meaningful relationships between words. Active layer 5) 50 units in the LSTM cell at this layer is to generate time dependence information in the text input sequence, which eventually helps the model learn complex contingent attribution within the sequence. The last Dense output layer using sigmoid activation function is because of binary classification tasks and its output is value of 0 to 1 as a probability. The model is compiled using Adam optimizer with learning rate 1e-4, to ensure a stable and fast learning. To do the testing, the give loss function is used which is binary cross-entropy, and it is widely used loss function for such binary classification problem and the model is evaluated based on accuracy. There are 9,686,151 trainable parameters to be trained.

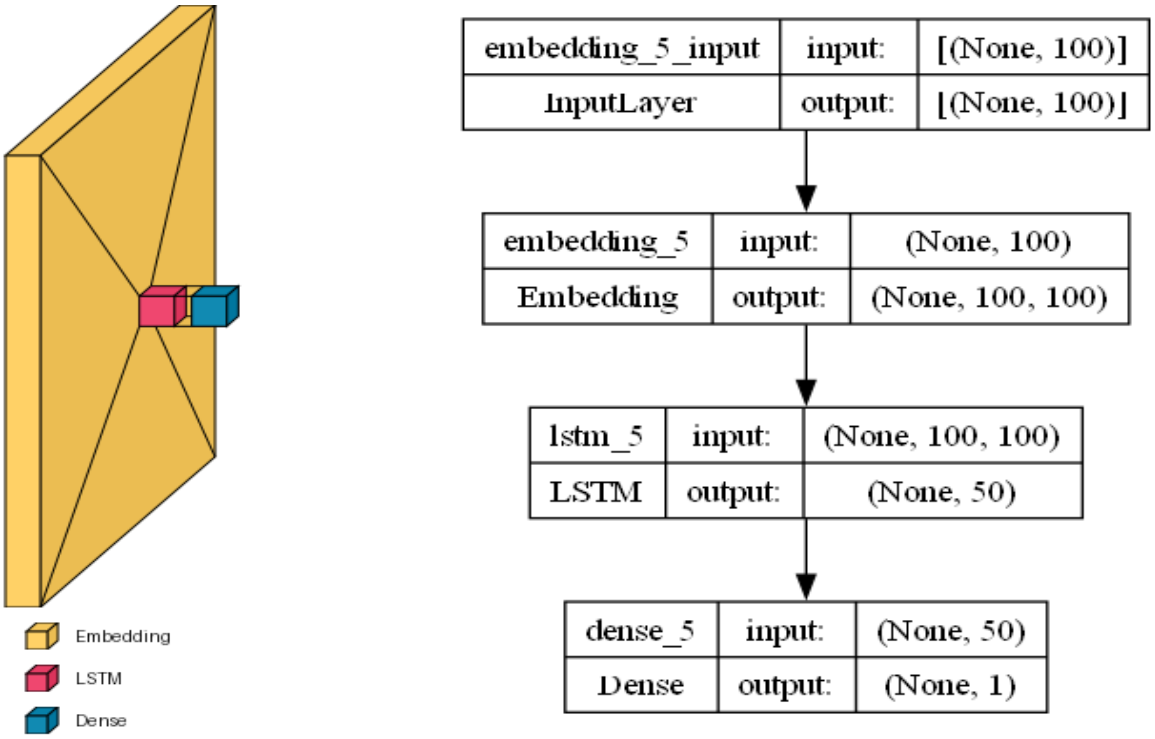

Figure 3.8: Architecture of Our Proposed LSTM Model

3.4.1.2 Gated Recurrent Unit (GRU)

The GRU presented in Figure 3.9 is a model of a recurrent neural network (RNN) that manages well some of the limitations of its earlier variant the RNN (here the long-term gradient decay problem) directly in virtue of its own way of handling long-term dependencies found in sequential data while learning. Closer to long short term memory (LSTM) networks, GRUs are costly in terms of computation because of the smaller number of parameters but they achieve similar performance in many tasks. There are two major parts of any GRU, these being the dependable update and reset gates, which decide what inputs to maintain or discard. It is a balance of what part of past information one should focus on and how much on the new data information, therefore, unifying the role of forget and input gates in LSTM. How much it forgets the past hidden while computing the new state is determined by the reset gate in this way the model can really take recent events into account. Unlike LSTMs, which maintain a single cell state for multiple time steps, GRU directly changes the hidden state, allowing for a simpler and faster network to be built. Because it doesn't have a separate cell state and updates the hidden state directly, it's, also, more minimalistic and interjection based. Hence, it is faster and needs less memory if it has less parameters for training and inference. Due to Then ability to encode long-term dependencies in a computationally feasible manner, GRUs proved to be a practical choice in several applications including speech recognition applications, machine translation, sentiment analysis, time series prediction. Because of its simplicity and efficiency, It is a viable replacement or alternative to LSTMs in complex or large scale models that require processing sequential data.

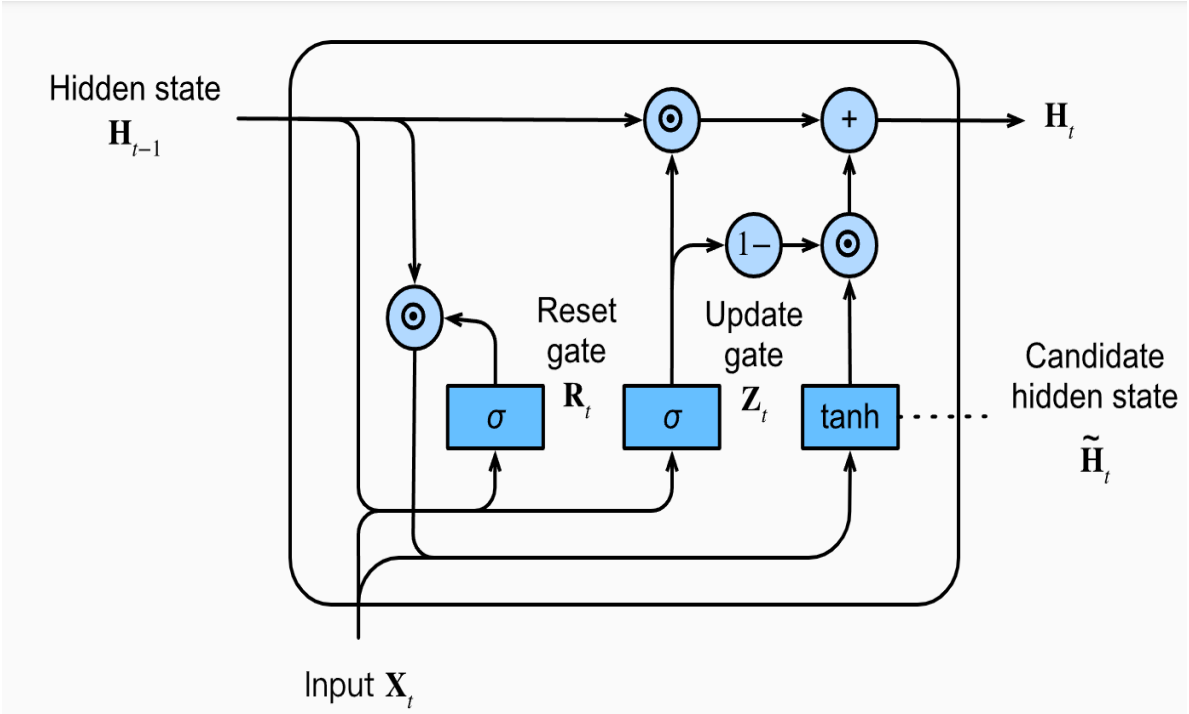

Figure 3.9: Architecture of GRU Layer

This is the sequential deep learning model.DSequential* for binary classification, which passes processed from the sequential data, as illustrated in Figure 3.10. Then further passed to an embedding layer, whose role is to convert words into general vectors or dense form of shape 100, such that the model could learn the semantics of this vectors. The max sequence length as input length in order to keep the input sequence lengths fix for uniform processing. Following a GRU layer with 50 units that model the long-term dependencies with less parameters than LSTM and hence better in terms of computation cost. GRU The GRU is similar to the LSTM since it has update and reset gates, however, it is relatively less complex and memory and computationally efficient than LSTM. Finally we have Denser output layer with sigmoid function as we are trying to find probability score between 0 and 1 for binary classification task such as spam detection, fake news classification, sentiment analysis etc. Connection uses an Adam optimizer learning rate of 10^{-4} to keep the learning stable and adaptive as it is effectively using the binary cross-entropy as its loss function which is standard for binary classification tasks. Its accuracy in prediction will be measured as the metric. There are 9,678,751 trainable parameters.

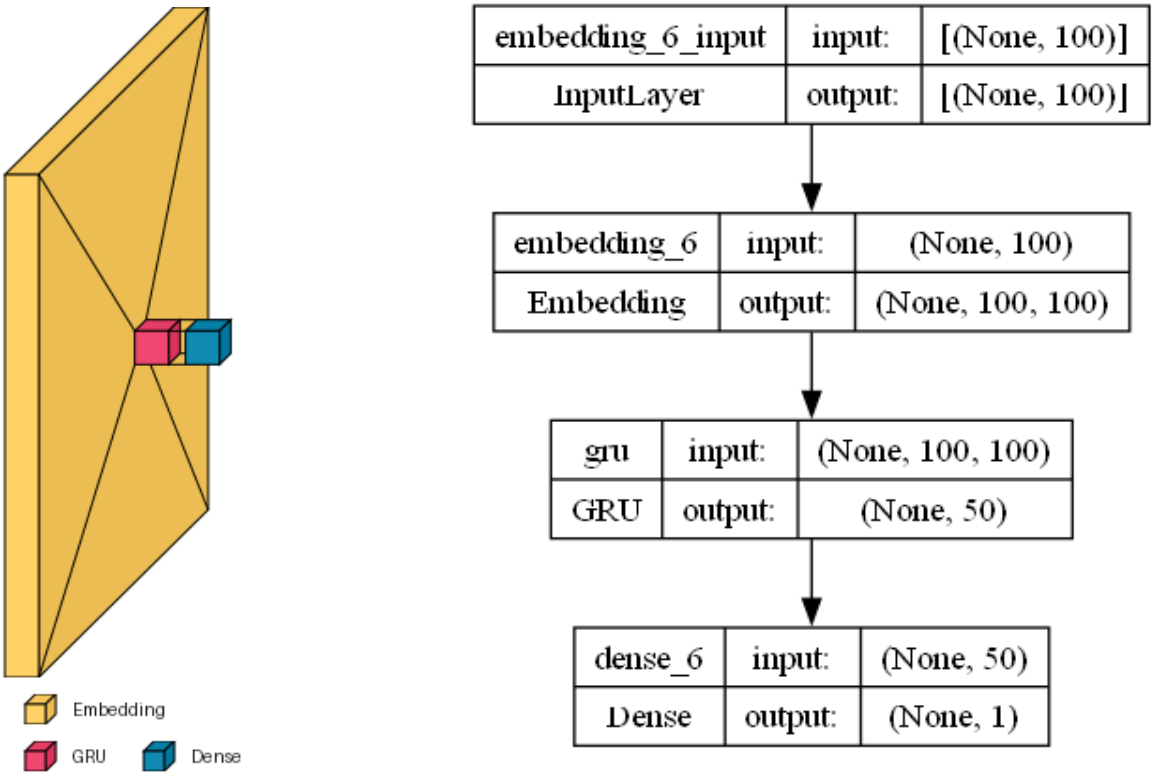

Figure 3.10: Architecture of Our Proposed GRU Model

3.4.1.3 Bidirectional LSTM (BiLSTM)

Figure 3.8 Modeling long-distance dependency with BiLSTM The BiLSTM The BiLSTM refers to the Bidirectional Long Short-Term Memory shown in Figure 3.11, which is a more sophiscated variant of the LSTM networks which are capable of further integrating past information with future dependencies within the sequences. Differently from an ordinary LSTM which reads one direction, the BiLSTM read the sentence in both directions simultaneously, using two distinct LSTM layers that one is in forward direction and the other is in backward direction. Thus, the model enriches representations of both their past and future context, making it efficient in dealing with complicated sequences. In its dispensation, the structure of the BiLSTM comprises two LSTM networks stored in cell states which have gate units that operate such as forget gate, input gate and output gate. The forget gate is laid out for what information to forget from the past, and input gate for how much information to add to the memory. Lastly, the output gate reads the cell state and from it decides what to output to the upcoming hidden state. Since BiLSTMs read data from both directions, they include more context, and to that extent can do better on a range of NLP, speech recognition, machine translation, ner and time series prediction tasks. However, BiLSTMs are computationally more complex as compared to LSTM because of their two-layer structure. There are in total 111111 trainable parameters.

3.4.1.3 Bidirectional LSTM (BiLSTM)

Figure 3.8 Modeling long-distance dependency with BiLSTM The BiLSTM The BiLSTM refers to the Bidirectional Long Short-Term Memory shown in Figure 3.11, which is a more sophiscated variant of the LSTM networks which are capable of further integrating past information with future dependencies within the sequences. Differently from an ordinary LSTM which reads one direction, the BiLSTM read the sentence in both directions simultaneously, using two distinct LSTM layers that one is in forward direction and the other is in backward direction. Thus, the model enriches representations of both their past

and future context, making it efficient in dealing with complicated sequences. In its dispensation, the structure of the BiLSTM comprises two LSTM networks stored in cell states which have gate units that operate such as forget gate, input gate and output gate. The forget gate is laid out for what information to forget from the past, and input gate for how much information to add to the memory. Lastly, the output gate reads the cell state and from it decides what to output to the upcoming hidden state. Since BiLSTMs read data from both directions, they include more context, and to that extent can do better on a range of NLP, speech recognition, machine translation, ner and time series prediction tasks. However, BiLSTMs are computationally more complex as compared to LSTM because of their two-layer structure. There are in total 111111 trainable parameters.

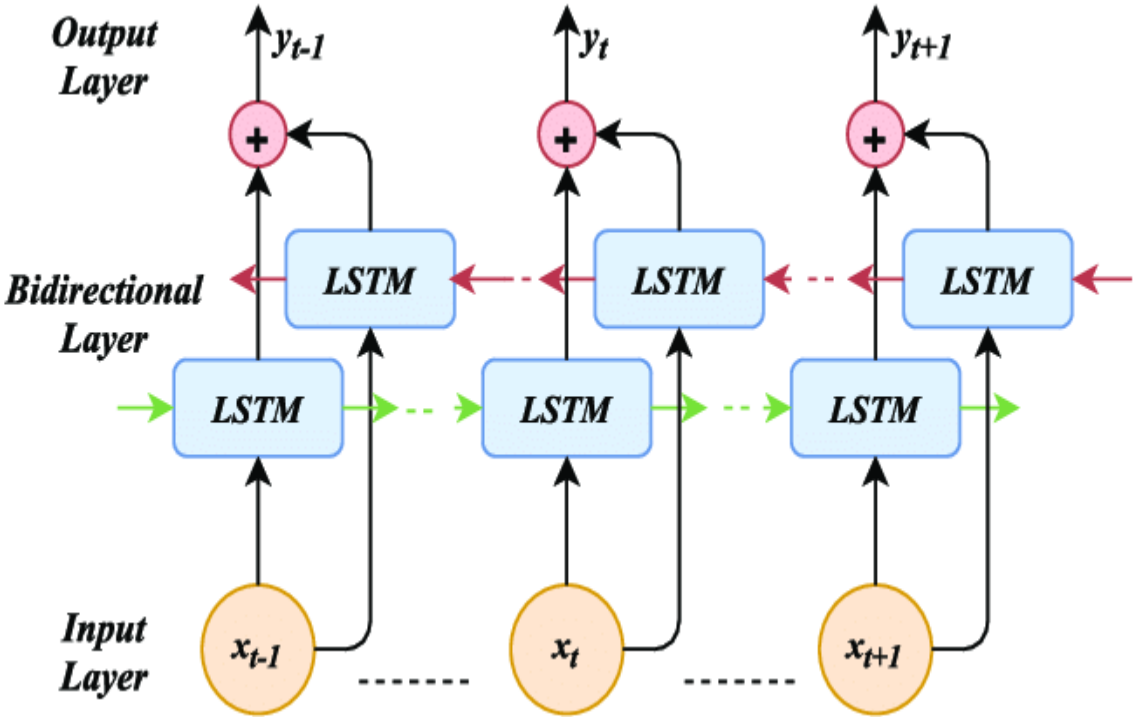

Figure 3.11: Architecture of BiLSTM Layer

The model of the proposition for the sequential prepares to performs binary classification applying a Bi-directional LSTM (BiLSTM) layer as illustrated in Figure 3.12, which is able to efficiently handle sequential data. It includes an Embedding layer which will turn positive integers into dense vectors of 100 dimension which could learn from and capture some semantic relationships between words. We set the length of the input to be max_sequence_length so all of the input sequences have the same length. The Bidirectional LSTM layer then has 50 more units and is essentially two LSTMs (one reads input forward, another reads input backward). The bidirectional format also allows the model to learn from both past and future context, strengthening the sequence model. The final layer is a Dense output layer using the sigmoid activation function in order to get a probability score in the range of 0 to 1 (which is the perfect choice for binary classification tasks such as sentiment analysis, spam detection, fake news classification). Trained using Adam optimizer with learning rate as 1e-4, the model receives smooth and effective updates of weights. We use a binary cross-entropy loss specifically aimed standard binary classification problems & accuracy as a main performance metric. The number of trainable parameters is 9,716,401.

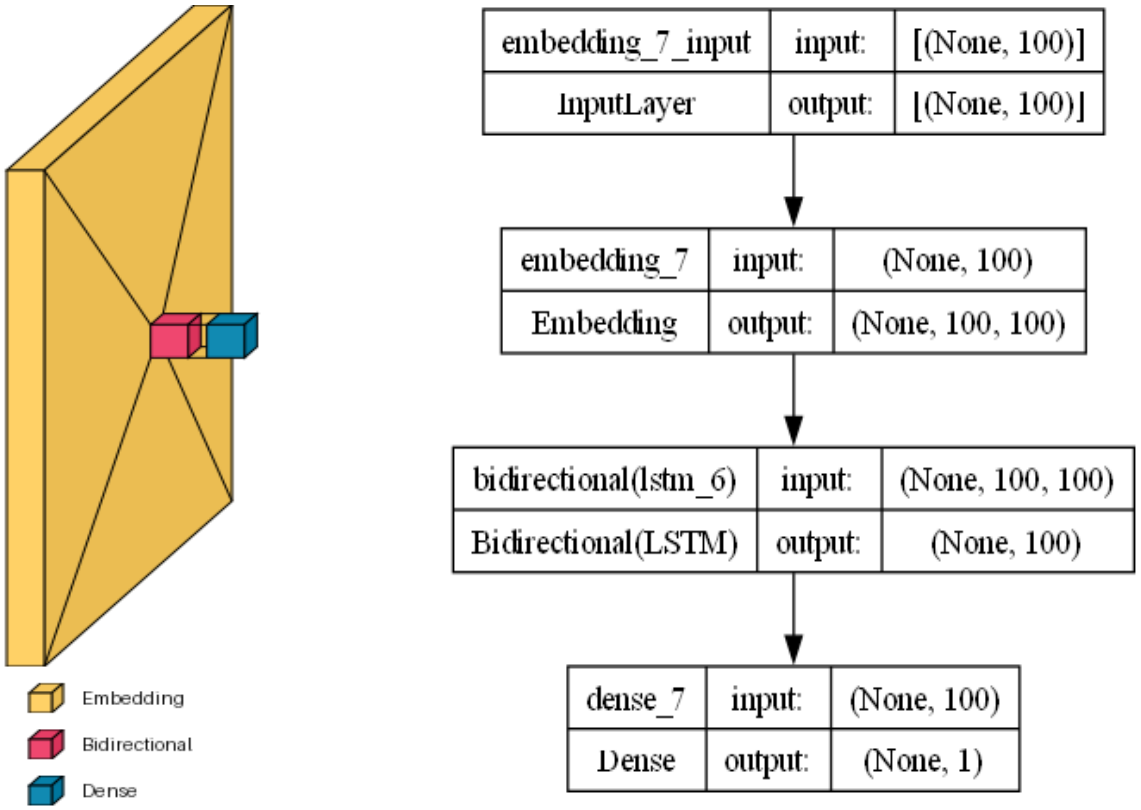

Figure 3.12: Architecture of Our Proposed BiLSTM Model

Bi-directional Gated Recurrent Unit (BGRU) shown in Figure 3.13, is advanced GRUs that can be used for sequence processing by remembering both past or future nodes. BGRUs differ from ordinary GRUs in that ordinary GRU layer processing is either forward or backward. A BiGRU has two GRU layers, one that moves forward and the other backward so that the model would learn richer representations of sequential data. It allows the network to maintain context on past and future inputs, making it excellent especially where more complex tasks relating to the sequence involved are concerned. The basic components of a GRU unit are an update gate and a reset gate that will manage the flow of information. An update gate determines how much past information is to be kept and changed with the new data, combining forgetting and input gates of LSTM into one mechanism. The reset gate controls the old information to be discarded while computing the new state; in this way, allowing the model to pay attention to more recent relevant information. Since BGRUs process text in both forward and backward directions, they offer better contextual learning than standard GRUs while holding computational efficiency. These are now commonly used for applications such as natural language processing (NLP), speech recognition, machine translations, sentiment analysis, and even time series. According to the research results, BiGRUs are faster and parameter efficient compared to BiLSTMs,resulting in their excellent performance in the same sequence modelling tasks.

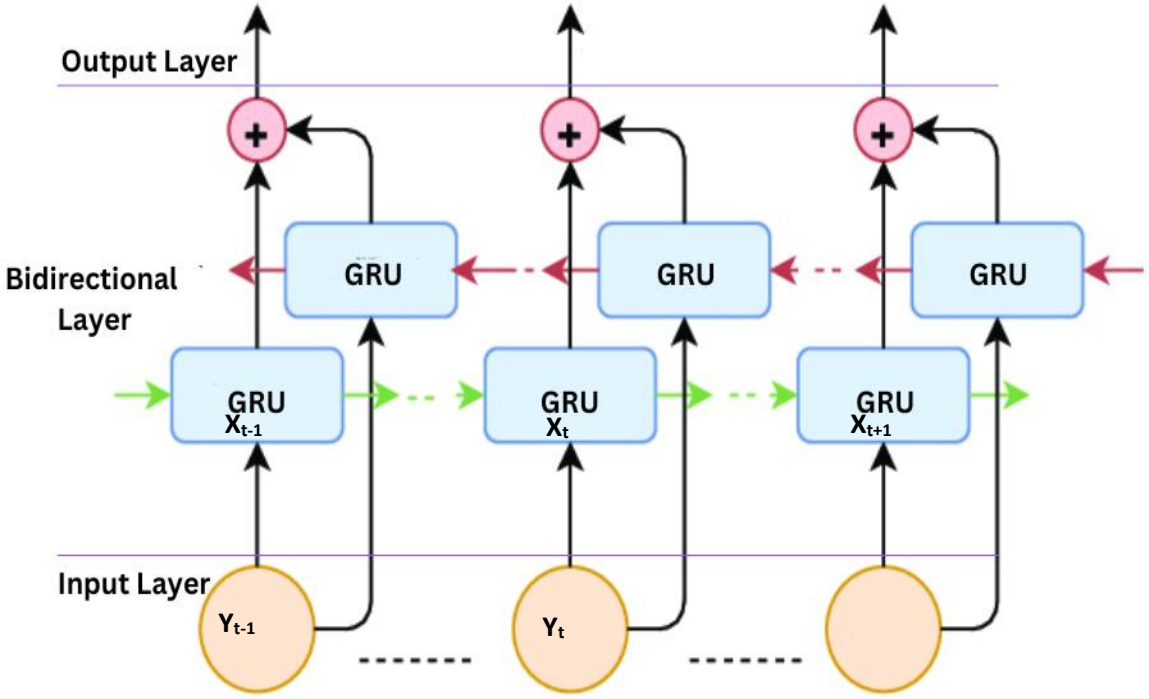

Figure 3.13: Architecture of BiGRU Layer

Binary classifier model that the model mentioned above is acutally using, which is a Sequential model, based on BiGRUs as depicted in Figure 3.14, TheCPC() improves the sequential data process effectively. First it works on an Embedding layer with words converting into dense 100-dimensional vectors. This enables the model to share semantic information across words. The input length is the maximum length of a sequence, which means all input sequences have the fixed length. The next is an bidirectional GRU layer with 50 units, which consistents of two GRUs which look over the input in forward, and backward directions. this architecture allows to the model to learn the repeated patterns both from the past and the future, leading to an enormous improvement in the way of evaluating the sequential patterns. Finally, we add a Dense output with a Sigmoid activation that produces a number between 0 and 1 which is just what we need to solve problems like spam detection/fake news classification/sentiment analysis. The model was compiled using Adam optimizer (with learning rate = 1e-4) for efficient and stable parameter updates. The loss function is binary cross-entropy, a common loss for binary classification tasks; the accuracy was adopted as the evaluation metric. Number of trainable parameters is 9,701,601.

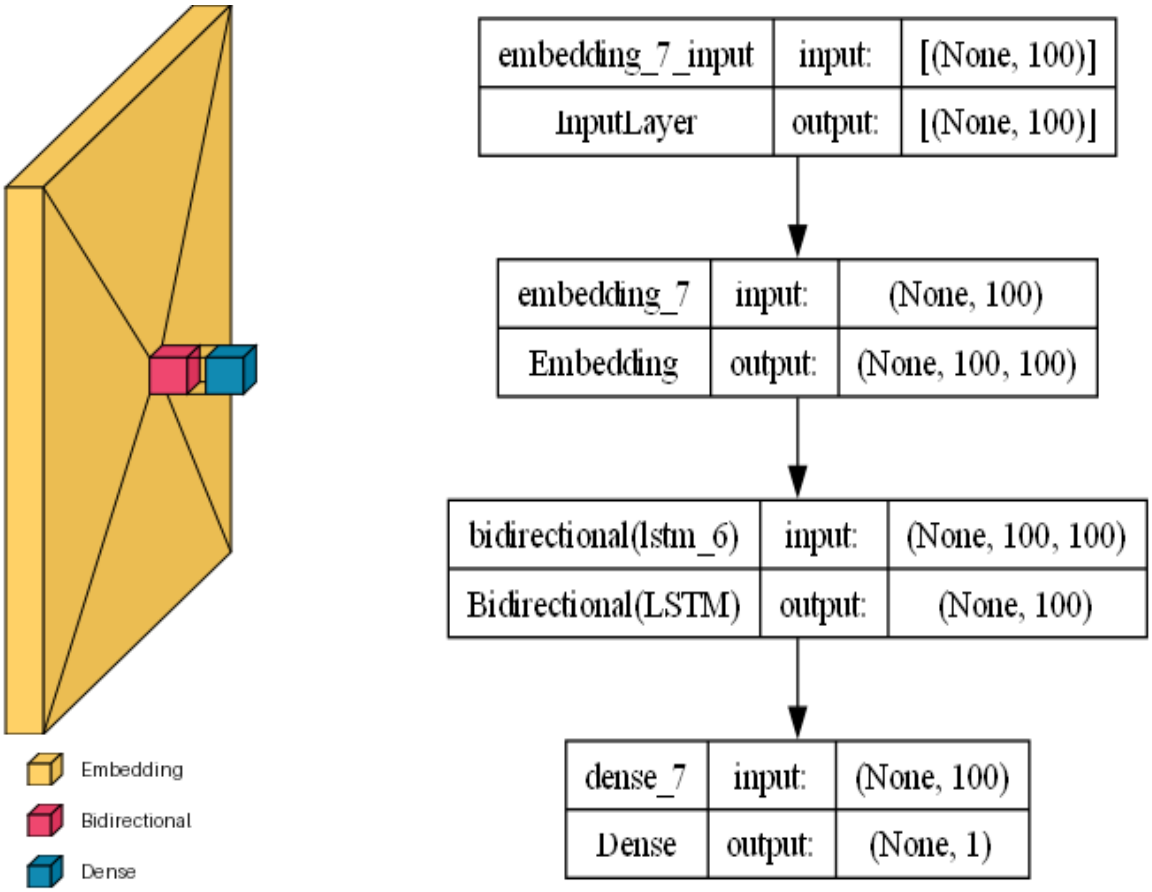

Figure 3.14: Architecture of Our Proposed BiGRU Model

3.4.2 Transformer-Based Model

3.4.2.1 Multilingual BERT

B Multilingual BERT (mBERT) is simple and effective extension of simple BERT for the multi-language understanding using a shared architecture. An extension to the base BERT (Bidirectional Encoder Representations from Transformers) model presented in Figure 3.15* BERT (Bidirectional Encoder Representations from Transformers) has been built by Google to handle over 100 languages with the architecture of the original BERT, the NLP model that revolutionized the way we learned context of words in a sentence using own self-attention mechanism model instead of a sequential approach (RNNs or LSTMs) that run through the words in order. Unlike the traditional pipeline processing paradigm, mBERT adopts the Transformer model, which supports parallel processing, resulting in substantial improvements in efficiency and accuracy in NLP. The architecture of the model is the same as BERT, and processing is performed by self attention and a feed forward neural network. It has two versions: BERT Base: 12 layers, 12 attention heads, 110 million parameters and BERT Large: 24 layers, 16 attention heads, 340 million parameters. Each transformer layer has a multi-head self-attention mechanism, a feed-forward neural network as well as layer normalization and residual connection which ensures the training stability and information from a previous layer be passed through the next layer. Pre-training itself is performed using the methods such as masked language modelling (MLM), in which some of the words in a sentence are masked and teacher signals for predicting them is provided, and next sentence prediction (NSP), where the model learns how the two sentences are connected.

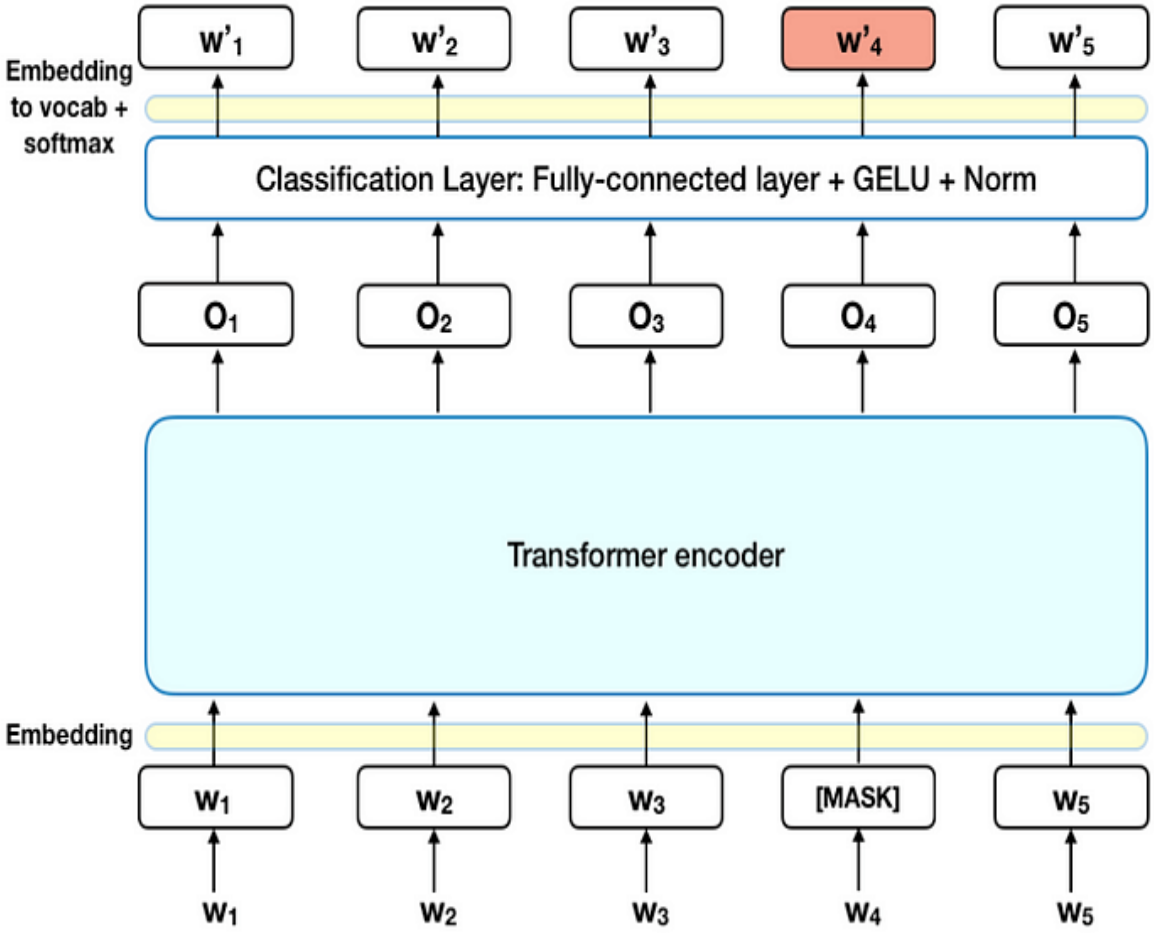

Figure 3.4.1: Architecture of BERT

The high proficiency of mBERT for Bangla fake news reactors is due to its capability to grasp the plethora of linguistic nuances and analyse them in deep context knowledge while trying to comprehend the misleading content and fake news. The task of Bangla fake news is challenging due to small labeled datasets, language characteristic, and highly code-mixed (Bangla and English) text. With 104 languages including Bangla in training, mBERT has exploited the potential of multilingual transfer and this enhances the performance even when language specific data is scarce. The model evaluates deception in the patterns of language, the relationships it creates in context and some of the tell-tales of misinformation, such as outrageous claims, sensational headlines and inconsistent narratives. Compared to traditional keyword-based approaches that typically adopt only sen- timent-agnostic specifications, mBERT's bidirectional attention mechanism enable to better infer the meaning of words in context, which can be more supportive for fake news classification. Moreover, mBERT is multilingual and can better understand code-mixed text, compared with monolingual models.

There are many cost advantages of mBERT with respect to fake news detection for Bangla. The multilingual pre-training leads the model for transferable knowledge for Bangla. By that virtue, mBERT can perform well even with a small fine-tuning data set. mBERT is unlike the rule-based systems in that it does not exploit a pre-defined complex pattern, instead it learns representations from the data to accommodate new patterns in the fake news in a constant manner. It's the rest of the NLP features that could be added to mBERT, such as Named Entity Recognition (NER) and Sentiment Analysis. Overall, mBERT largely blunts new transition lines against misinformation in the Bangla media. It is an extremely scalable and effective method that people would use to detect the spread of fake news. Here, we found mBERT to be awesome when it comes to analysing and counteracting the knife-stabs of misinformation in the Bangla digital arena, bearing the transformer

driven architecture, multilingual learning and the deep context understanding capacity.

3.5 Model Training

3.5.1The training of the deep learning model

(22) The deep learning models were trained using the Adam optimizer with a learning rate of 1e-4 and binary-cross entropy loss for the binary classification. The evaluation metric chosen was accuracy. They ran with a batch size of 32 for a total of 10 training epochs, as well as with a 20% validation split to observe performance as an anti-overfitting measure.

3.5.2 Transformer Based Model Training

Fine-tuning was conducted for 10 epochs with batch size of 16, and learning rate of 2e-5 using mBERT (bert-base-multilingual-cased) models. A weight decay of 0.01 was used for further generalization. The training also included epoch-wise evaluation, logging, and checkpoint saved with an accuracy metrics for the best model selection. The length Training argument for this model is given in Table 3.1.

TABLE 3.1: TRAINING ARGUMENTS FOR MBERT MODELS

HYPERPARAMETER	VALUES
TRAINING EPOCHS	10
BATCH SIZE	16
LEARNING RATE	2e-5
WEIGHT DECAY	0.01
EVALUATION STRATEGY	epoch
LOGGING STRATEGY	epoch
SAVE STRATEGY	epoch
BEST MODEL METRIC	accuracy

3.6 Model Evaluation

Finally, the model was applied in a testing set for generalization testing. The second objective here was to calculate the loss and accuracy metrices to verify how good a model is at classifying new data. Moreover, precision, recall and F1 were also estimated to verify the results of the model. The formulae for precision, recall, F1, and accuracy are given below.

$$Precision = \frac{TP}{TP + FP} \dots\dots\dots\dots\dots\dots\dots (1)$$

$$Recall = \frac{TP}{TP + FN} \dots\dots\dots\dots\dots\dots\dots\dots (2)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \dots\dots\dots\dots\dots\dots\dots (3)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \dots\dots (4)$$

Here, TP refers for True Positive, FP refers as False Positive, TN refers as True Negative

and FN refers as False Negative.

3.7 Project Plan

The detailed project plan shown in Figure 3.7.1

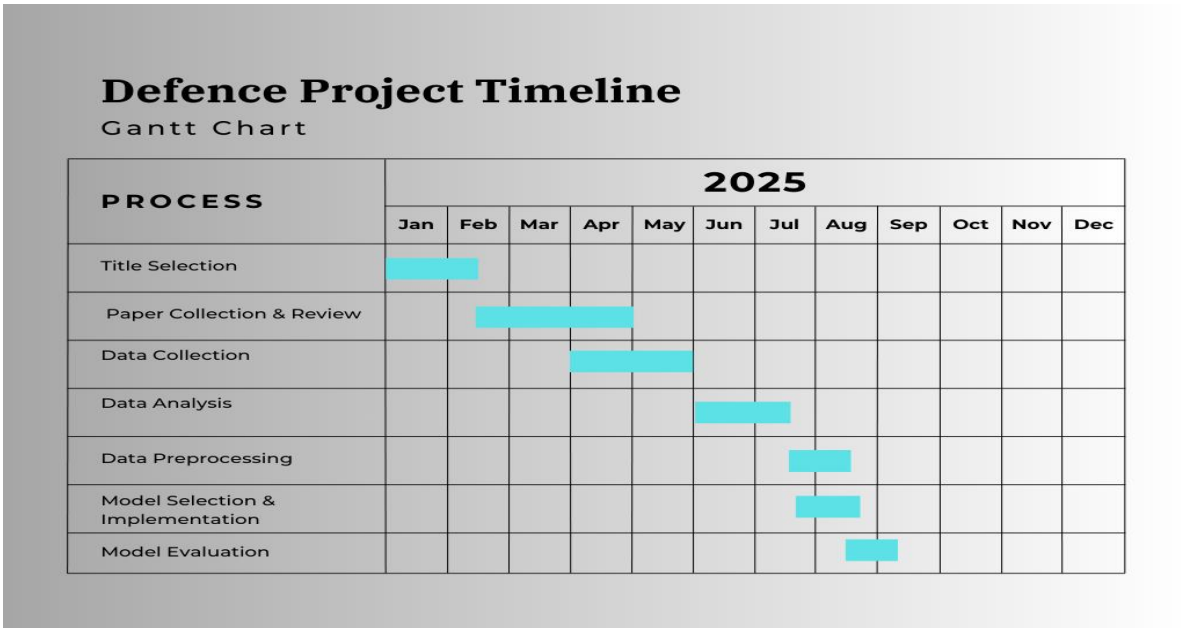

Figure 3.7.1 Project Plan

Chapter 4

RESULT AND ANALYSIS

4.1 Overview

This section demonstrates the operational performances of the Bangla fake news detection and analysis based on significant metrics of accuracy, precision, recall, F1-score, and loss curves that have the ability to develop for the behavior, the model during training, and validation. Comparison between architectures Comparing different architectures to shows the performance of models in specific settings and highlights its responsible strengths and weaknesses. Results may be used to draw conclusions about generalizations of the model, overfitting mechanics and ways to reduce classification mistakes.

4.2 Evaluation of the Deep Learning Based Model

4.2.1 LSTM Model Evaluation

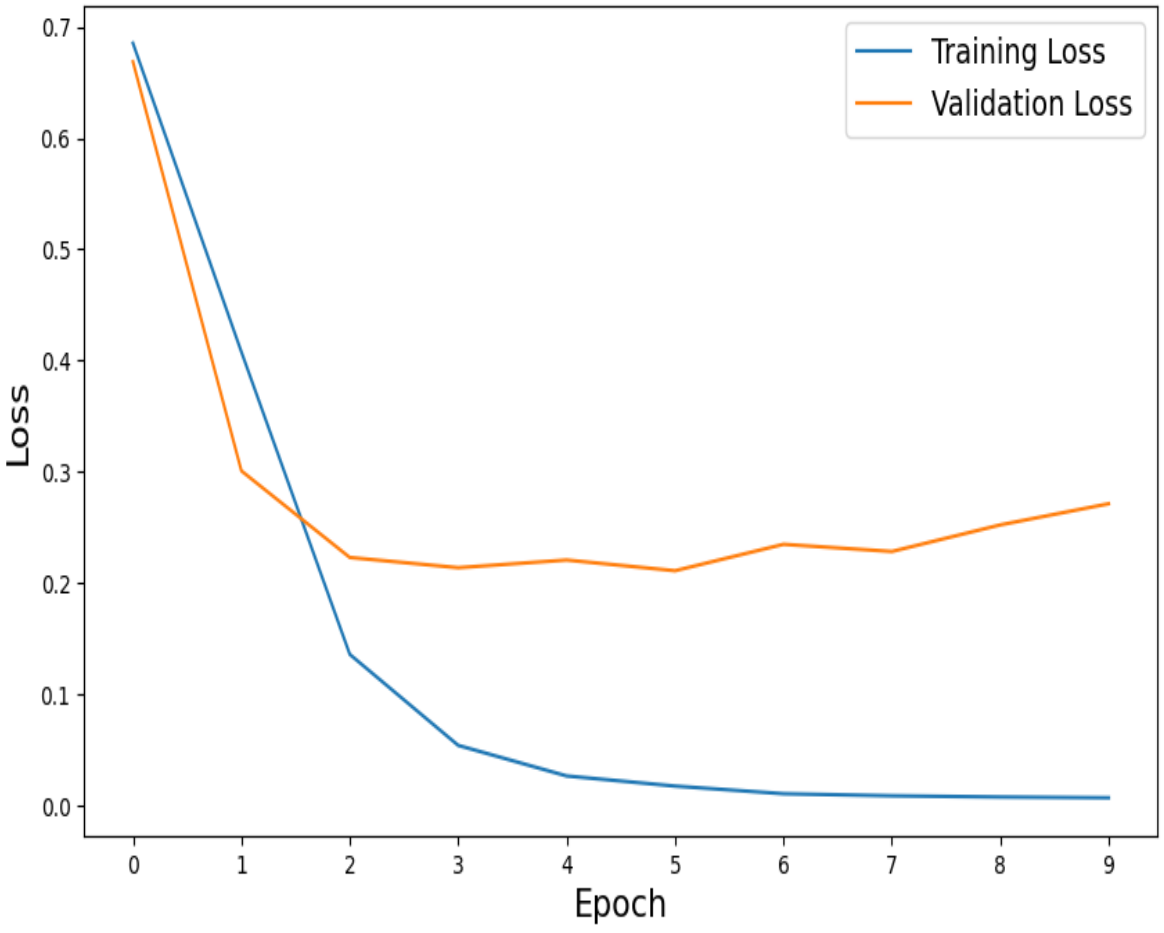

Figure 4.2.1: Training vs Validation Loss Curve of the LSTM Model

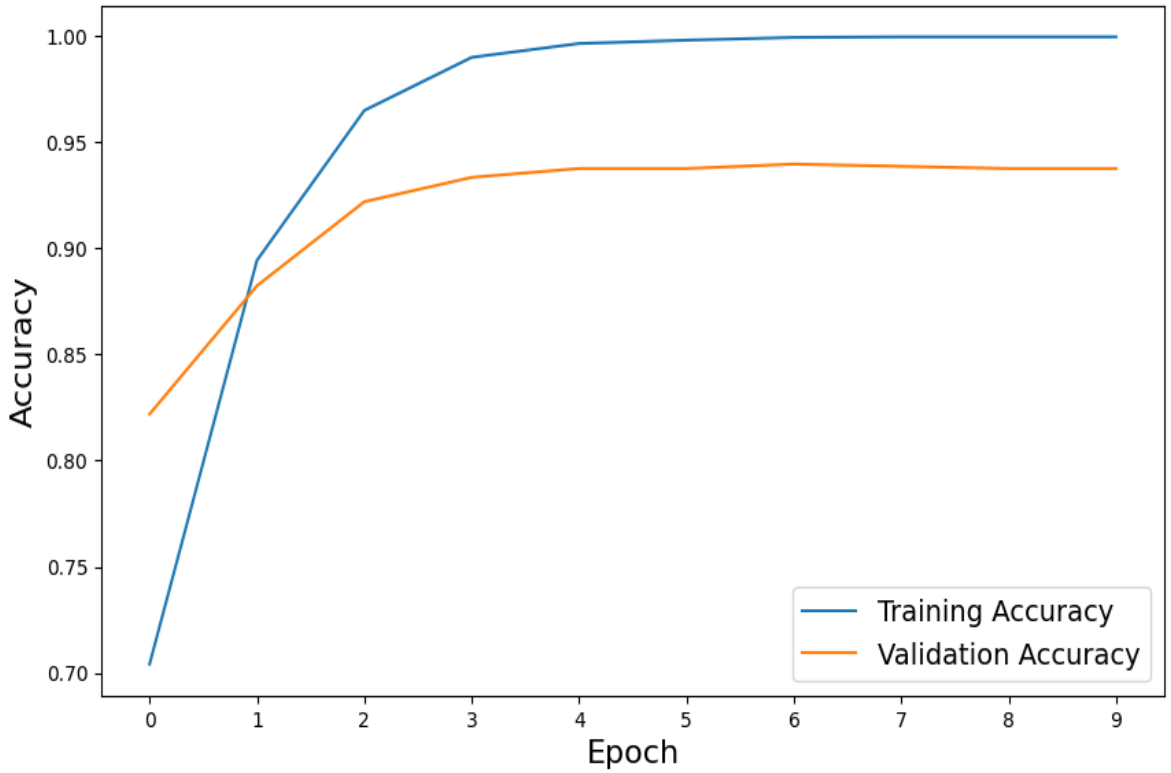

Figure 4.2.2: Training vs Validation Accuracy Curve of the LSTM Model

The learning curves of training and validation loss of the LSTM model can be seen in Figure 4.1. The train loss is gradually decreased as the network learnt the dataset properly. The decrease in the validation loss slows down and after epoch 3 it pretty much stops and even increases slightly. This is an indication of overfitting, i.e., the model is beginning to memorize the training examples rather than generalize well to unseen data. The gap between training and validation loss begins to widen which indicates that overfitting begins to occur and that regularization techniques such as dropout or early stopping may be used to prevent overfitting and increase model generalization.
Akin to Figure 4.1, Figure 4.2 shows the curves for the training and validation accuracies. Training accuracy ramps up quite fast, close to 100% within a couple epochs, but validation accuracy gets better and then flattens out at approx 92-93%. The difference between training and testing accuracy also confirms that overfitting occurs. The validation accuracy curve is close to horizontal for the latter epochs, indicating that, although the model may incorporate some useful features, the level to which it can generalize is seemingly not getting better beyond that point. If that were the case the the more you tune hyperparameters, the larger amount of diversity you add to the dataset, the higher incorporation of extra augmentation steps, the better the model preform. It demonstrates good performance on classification level over the barriers; thus is a promising model for text classification tasks.

TABLE 4.1: CLASSIFICATION REPORT OF LSTM MODEL

CLASS	Precision	Recall	F1-Score
0(FAKE)	0.94	0.93	0.94
1(REAL)	0.94	0.95	0.94
ACCURACY			0.94

MACRO AVG	0.94	0.94	0.94
WEIGHTED AVG	0.94	0.94	0.94

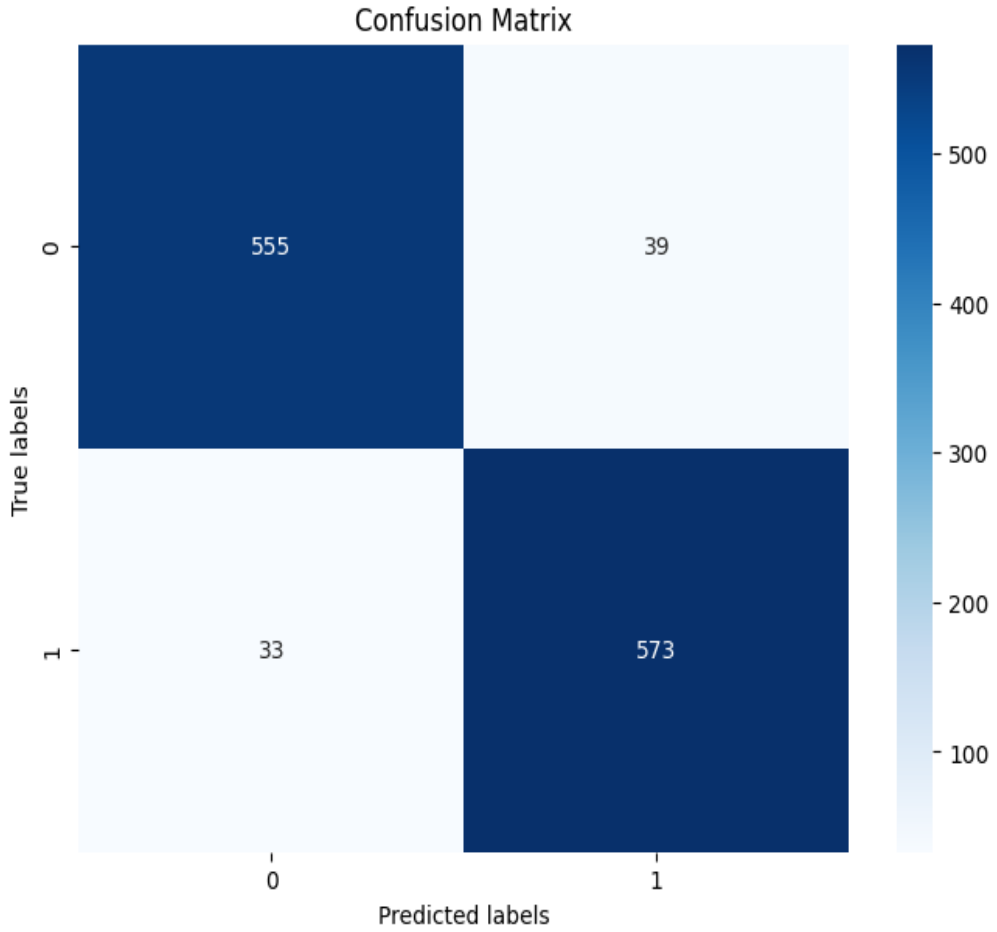

Figure 4.2.3: Confusion Matrix of the LSTM Model

The classification report for the LSTM in Table 4.1 had a precision of 94%, indicating that the LSTM did an excellent job. We have high value precision and recall for two classes (real and fake news) The precision for real news is 0.94 and for fake news is 0.94. The recall for real news is 0.93 and for fake news is 0.95. A 0.94 F1-score for both real and fake means that they are able to identify both real and fake news with little to no bias. The model is also rated at 0.94 in terms of both macro average and weighted average for precision, recall, and F1-score, resulting in the model's stable classification. Therefore, it is also confirmed that the classification between genuine news and fake news is best backed by the LSTM model. Nevertheless, enhancement is within reach by further optimizing the model with more regularization strategies or by leveraging more various datasets.

The confusion matrix plotted in Figure 4.3 sheds further light onto the model's classification decision. The classifier got 555 real news as true positive and 573 fake news as true negatives; for the false positives, the real were classified as fake - negative (39 samples) and for the false negative, the fake was classified as real (33 samples). Rates of false positives and negatives are low, meaning the model correctly predicts all else, however there remains a healthy amount of room for improvement. Further error reduction might be obtained through class weighting, expanding the training set to a more diverse set of data or ensemble learning approaches. Despite that minor misclassification, the output of the LSTM model would clearly separate real news from fake news, which therefore made it very effective to combat fake news.

4.2.2 GRU Model Evaluation

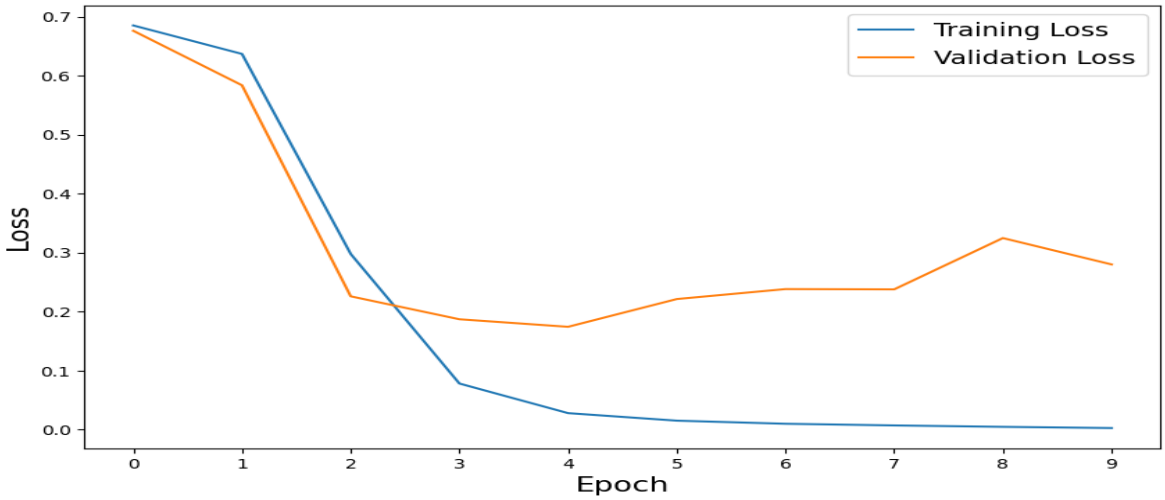

Figure 4.2.4: Training vs Validation Loss Curve of the GRU Model

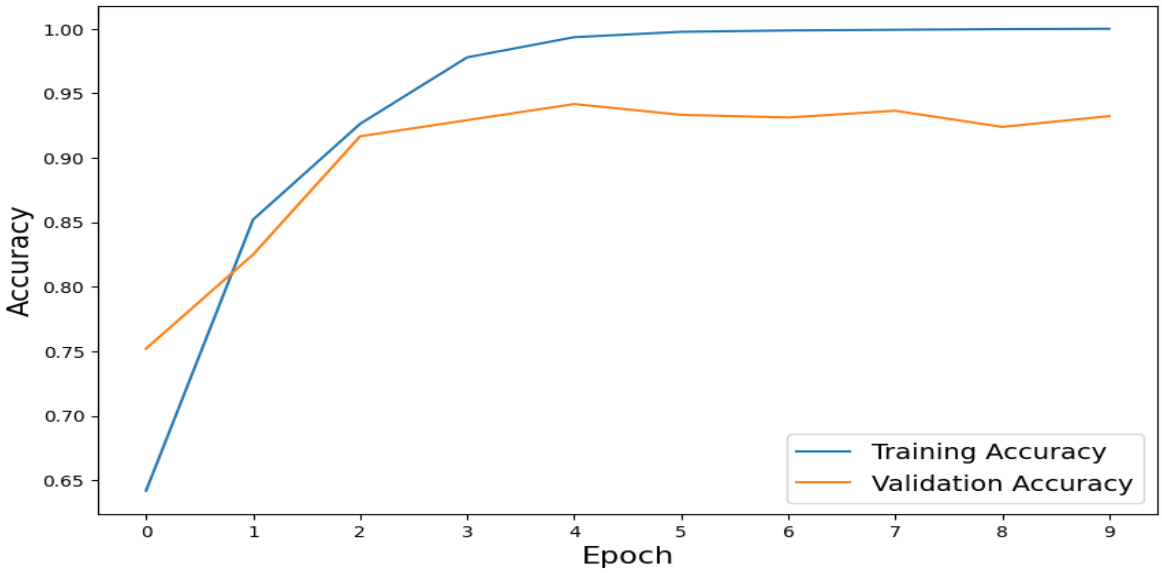

Figure 4.2.5: Training vs Validation Accuracy Curve of the GRU Model

The training and validation loss curves of the GRU model are depicted in Figure 4.4. The training loss keeps decreasing, we can thus see that the model is capable to learn patterns in the train data. The validation loss initially decreases minimally, which means that good generalization is achieved in the first epochs, and then fluctuates and diverges, indicating overfitting. This kind of behaviour does suggest that although the model has to some extent learned good features of the training data, in further epochs it has started to lose some of its capability of generalization on the unseen data. Dropout's regularization, early stopping or adapting learning rates can help to mitigate this issue and improve generalization of the model.
Figure 4.5 as the curves for training and validation accuracy. CNN learns very well on training data (close to 100% after first few epoch) and we achieve validation accuracy of 93-94%. The increasing difference between the training and validation loss gives more proof for overfitting since the model trains well on training data but does not generalize well on validation data. However, validation accuracy is still high, which means that GRU can still classify fake vs real news with very good results. Fine-tuning the model could improve its generalization, e.g., by reducing model complexity or using more elaborate regularization.

TABLE 4.2: CLASSIFICATION REPORT OF GRU MODEL

CLASS	Precision	Recall	F1-Score
0(FAKE)	0.94	0.94	0.94
1(REAL)	0.94	0.94	0.94
ACCURACY			0.94
MACRO AVG	0.94	0.94	0.94
WEIGHTED AVG	0.94	0.94	0.94

The classification report of the GRU model in Table 4.2 shows that in general, there is a good accuracy, 94%, that represents its capability in separating the fake news from the true ones. For all classes, precision, recall and F-1 score 0.94 are reported, indicating a fair classification task. The macro and weighted averages for all metrics are also 0.94, which implies that the model performs consistently over the both classes. Thus, it means that the GRU model is able to in general detect fake news well without the severe class imbalance. The generalization of the model could be improved even further through some hyperparameter optimization or more training data included, or more advanced form of regularization.

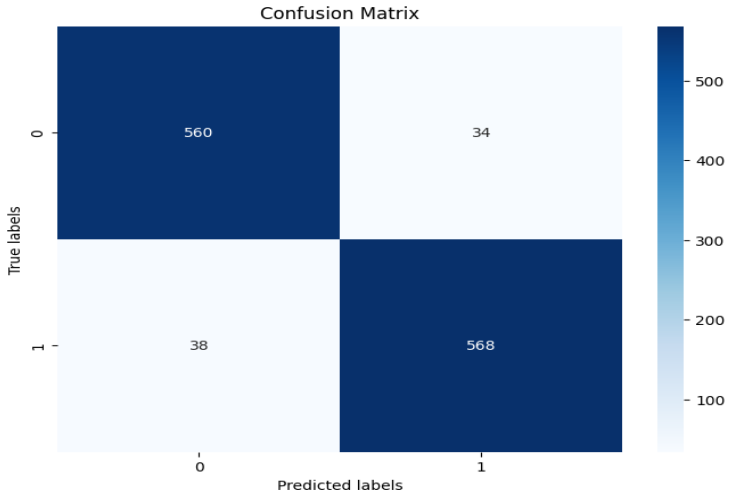

Figure 4.2.6: Confusion Matrix of the GRU Model

Figure 4.6 represents a confusion matrix indicating decisions made by the model in a more tangible way. The model predicts 560 real news samples as real and 568 fake news samples as fake. False positive stories consist of 34 which misclassify real news as fake. False negatives consist of 38 reports that fake news is real. Error on rates under misclassification demonstrate a realizable low number for the GRU model only when it learns from unseen data. But it is the issue of false negative, which demands more attention in news detection, as a skewed fake masquerading as real would result in catastrophic consequences. Enhance classification: (#classes: 13) may be accomplished by ensemble modelling, fine tuning learning rate, or using domain-specific word embeddings. With those marginal misclassifications aside, GRU was predictable and hence, a good opponent for fake news.

4.2.3 BiLSTM Model Evaluation

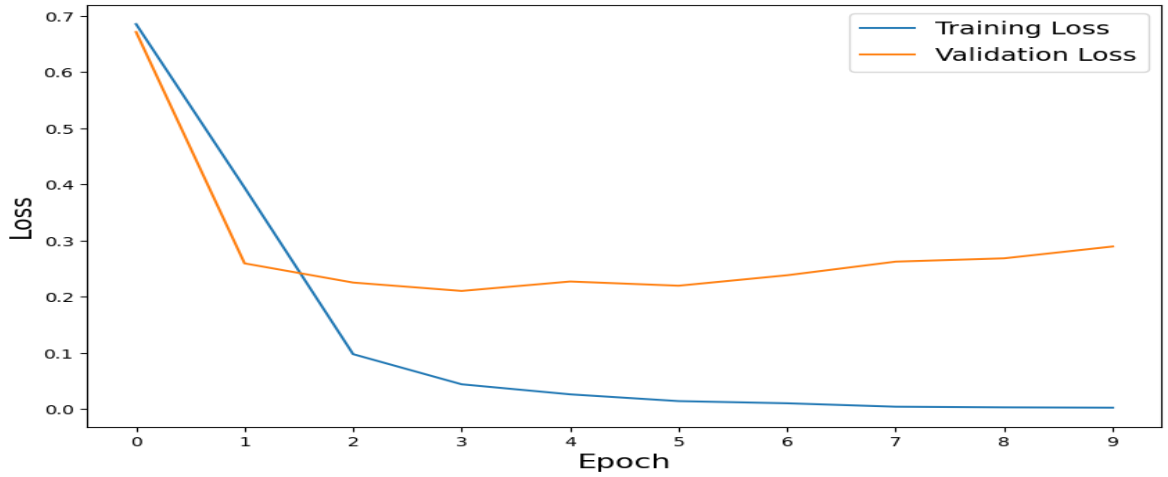

Figure 4.2.7: Training vs Validation Loss of the BiLSTM Model

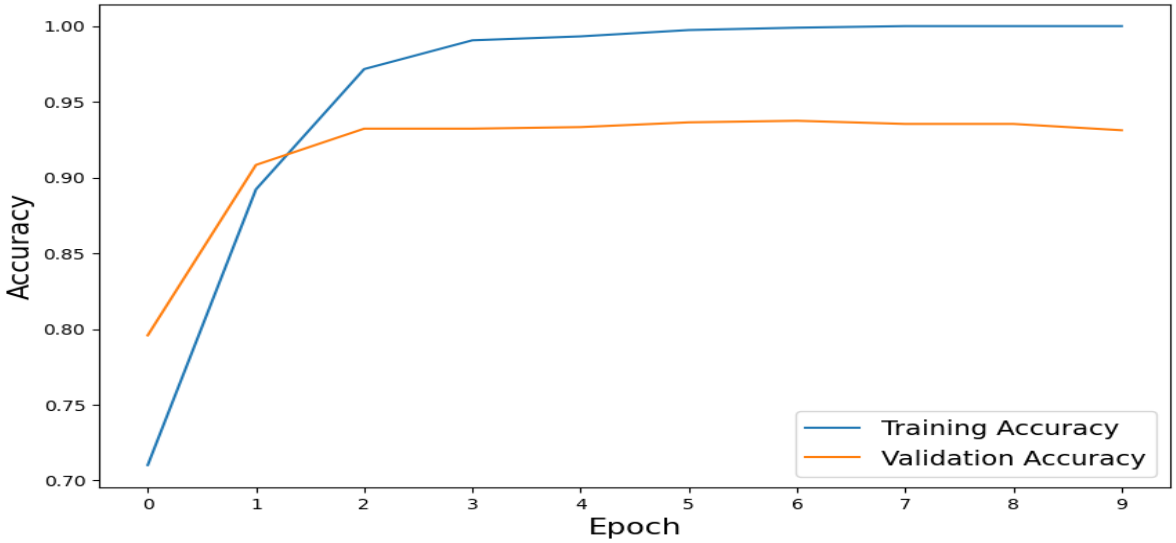

Figure 4.2.8: Training vs Validation Accuracy of the BiLSTM Model

As depicted in Figure 4.7, the training and validation loss curves for the BiLSTM model reveal that towards the end of the training session, the training loss keeps decreasing, whereas the validation loss has its initial decrease for a few epochs and gets plateaued before rising. This trend indicates that the model is able to represent the data well at the beginning of the training but it begins to overfit over time. Furthermore, and this growing gap shows then that both dropout and weight decay or possibly early stopping are necessary techniques to be used in order to facilitate generalization.

Figure 4.8 shows the training and validation accuracy curves. The training accuracy quickly learns to nearly 100%. The validation accuracy, however, is constant around 93-94% values. This kind of results demonstrates that the BiLSTM model can learn relatively easily from training data, but not so well from unseen data (in validation).

The constant posterior epoch validation accuracy means that model has absorbed useful features and it might need just some fine-tuning to reach the performance. But the performance of the BiLSTM model is promising in the classification task and can be a reliable method for Bangla fake news classification.

TABLE 4.3: CLASSIFICATION REPORT OF BILSTM MODEL

CLASS	Precision	Recall	F1-Score
0(FAKE)	0.93	0.94	0.94
1(REAL)	0.94	0.94	0.94
ACCURACY			0.94
MACRO AVG	0.94	0.94	0.94
WEIGHTED AVG	0.94	0.94	0.94

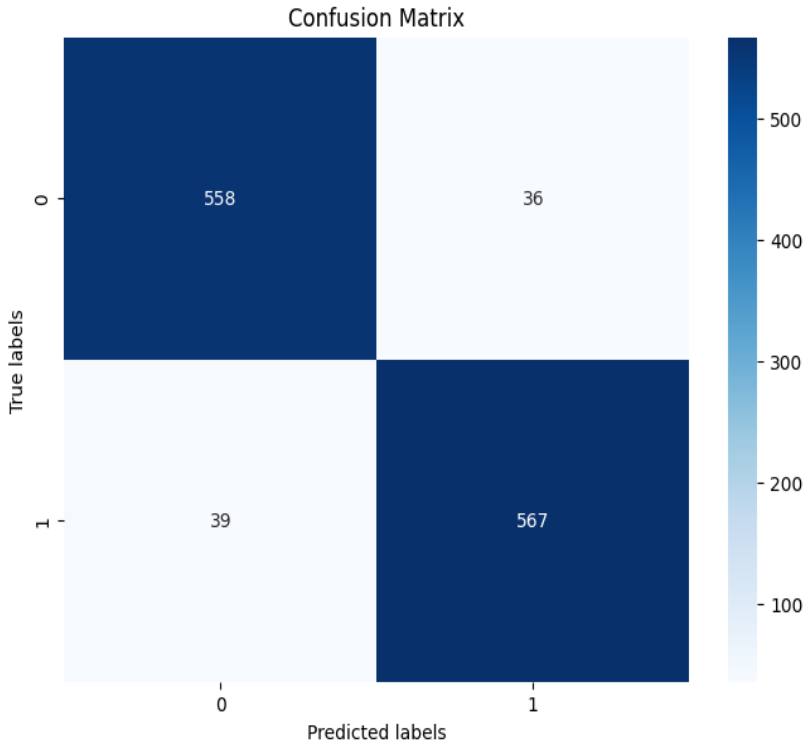

Figure 4.2.9: Confusion Matrix of the BiLSTM Model

The learned results of BiLSTM on this dataset was further validated with the classification report in Table 4.3. We obtain an overall accuracy of 94% using the BiLSTM. The precision was 0.93 for real news and 0.94 for fake news, indicating the model successfully minimized false positives. Both classes get roughly the same recall score (0.94), which explains that the RoBERTa is as good at classifying fake and real news correctly. Moreover, F1 score for both classes is again considerably high at 0.94, proving the BiLSTM model takes a balanced position between precision and recall. Additionally, macro and weight mean averages are near-constant at 0.94, validating that the model is appropriately calibrated across varied cross-categories. However, as results are high accurate, little changes such as the tuning of hyper-parameters or the use of task specific embeddings might

lead to an enhancement of the performance.

The confusion matrix is shown in Figure 4.9 that gives a thorough examination of BiLSTM model's classification capability. The model correctly predicted 558 true news and 567 false news, but also misclassified 36 true news samples as false and 39 false news samples as true. Although the magnitudes were low, to the best of our knowledge, reducing false negatives is crucial for their application as the mislabeling of fake news as real can be leveraged to spread false information. Further improvements can include adjusting the learning rate, using attention mechanism or ensemble methods for high precision. As the number of errors is very low, the BiLSTM classifier is still one of the most effective methods for text classification with a significant track record.

4.2.4 BiGRU Model Evaluation

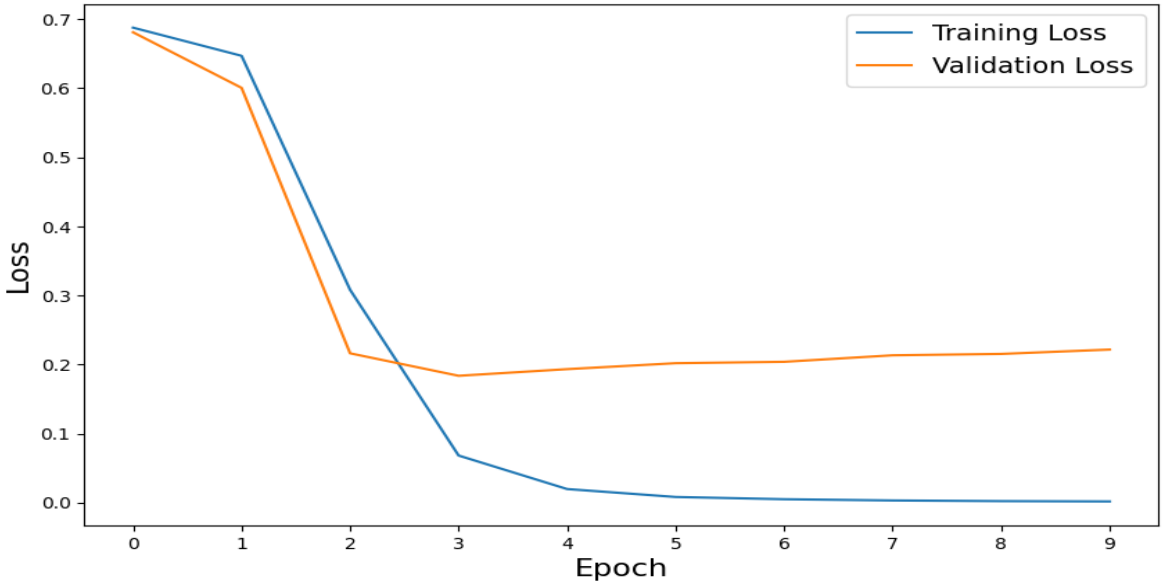

Figure 4.2.10: Training vs Validation Loss of the BiGRU Model

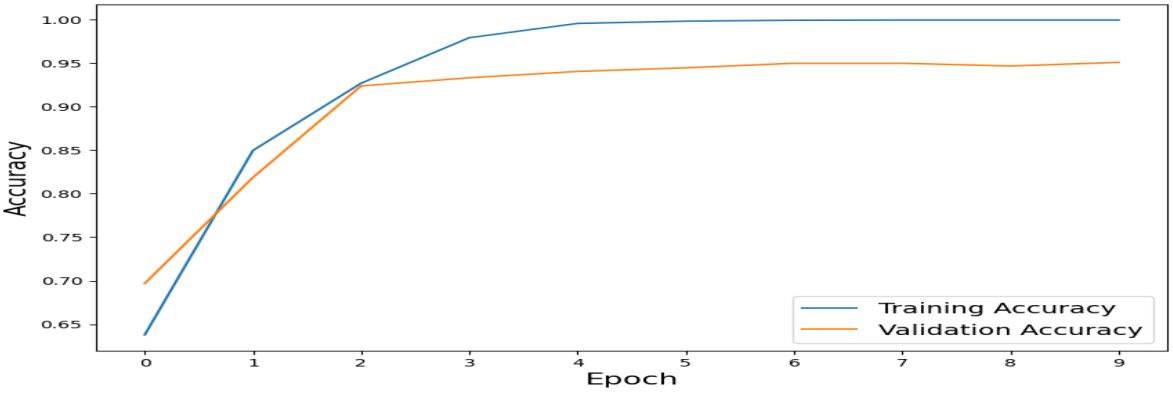

Figure 4.2.11: Training vs Validation Accuracy of the BiGRU Model

Figure 4.2.10, which show the loss curves of training and validation for the BiGRU model. The value of the training loss is also continuously decreasing, and the values are changing well by learning in a dataset. It's just that now the validation loss is actually going down, but it starts showing signs of plateauing and a possibility of slight increase after a few epochs —overfitting much? The

increasing values of the training and validation loss indicates the need of regularisation tools such as dropout or early stopping in order to enhance the model's generalisation. Nevertheless, the validation loss of this model is fairly low, suggesting good prediction ability.

TABLE 4.4: CLASSIFICATION REPORT OF BIGRU MODEL

CLASS	Precision	Recall	F1-Score
0(FAKE)	0.94	0.97	0.95
1(REAL)	0.97	0.94	0.95
ACCURACY			0.95
MACRO AVG	0.95	0.95	0.95
WEIGHTED AVG	0.95	0.95	0.95

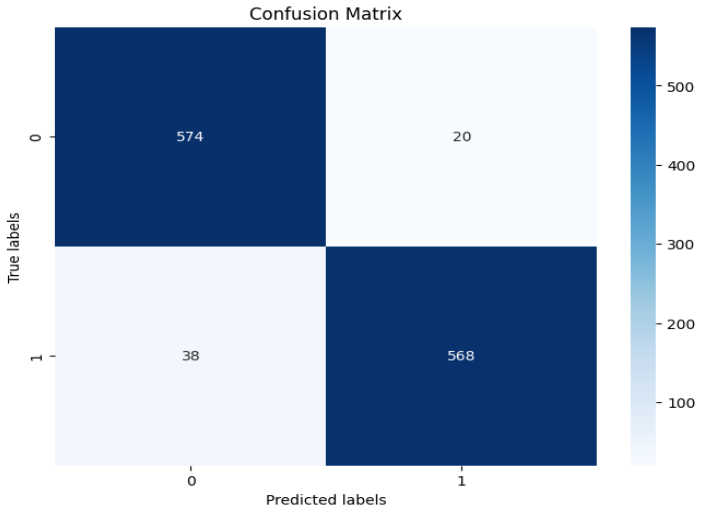

Figure 4.2.12: Confusion Matrix of the BiGRU Model

The BiGRU model in table 4.4's classification report demonstrates that it has significant performance on overall accuracy, achieving 95% overall accuracy which outperforms the compared models. Model precision is 0.94 for real news and 0.97 for fake news, which showed that false positives were reduced efficiently. The recalls are 0.97 and 0.94 for real news vs. fake news respectively, which indicates that performance of real news is slightly higher than the other one but has a correct level of real news. The balanced F1-score of 0.95 for both of the classes proves the fact that the BiGRU model has good trade-off between precision and recall. Furthermore, the macro and weighted average are all 0.95, which means that the model has consistently good performance from two aspects, and also shows that the BiGRU is an effective model in text classification, especially for fake news detection.

Figure 4.12 provides a confusion matrix to provide a further insight into the classification performance of the model. For real news the BiGRU model correctly classifies 574 samples, for fake news, it correctly classifies 568, and the number of false positive (i.e. real news which are mis-classified as fake) is 20, likewise false negatives (i.e. fake news which are mis-classified as real) is 38. It suggests a better model sign, since the quantity of false positives is quite lower than the rest, which

means BiGRU has good performance in real news removal. The false negatives goes slightly high, that means some pieces of fake news are detected as true. Work There To improve further the accuracy, we could apply hyperparameter tuning or use other methods like the attention mechanism or ensemble learning. Despite these few mistakes, the classification performance of the BiGRU model is robust and it can be regarded as being a good candidate for real-world use of fake news detection.

4.3 Transformer Based Model Evaluation

4.3.1 Multilingual BERT Model Evaluation

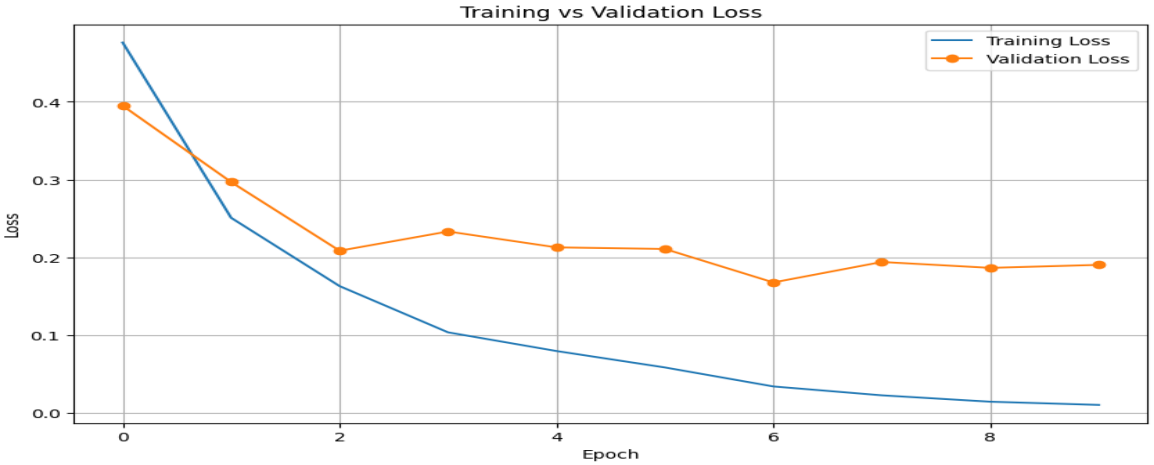

Figure 4.3.1: Training vs Validation Loss of the mBERT Model

In Figure 9 we can see that the loss curves of training and validation are performed on the mBERT model. The value of the training loss keeps decreasing which indicates good learning with convergence for the model. The validation loss kept falling during the first epochs, then converged quickly with only marginal oscillation, on a range that possibly covered the learning of relevant patterns by the model, followed by no significant improvement in the performance. This small increase in validation loss at the end epochs is suggestive of overfitting happening with mBERT model but much less as compared to a recurrent model. This suggests that the architecture design of the transform-based model that used in mBERT may enable a better generalization, while other training parameters tuning including learning rate, or using an early stopping could used in considering the performance.

In contrast to traditional models, mBERT is built on self-attention mechanisms from transformers which allow the model to preserve long-range dependencies on textual data. Despite some variations in validation loss, the model prediction is significant supported by the low-loss values. A lower distance between training and validation loss directly shows that less overfitting threat is carried by mBERT than LSTM or GRU elucidating superiority in terms of generalization. The further fine-tuning, or training with more data, or using domain specific pre-trained embedding will definitely boost any other classification accuracies and robustness.

TABLE 4.5: CLASSIFICATION REPORT OF MBERT MODEL

CLASS	Precision	Recall	F1-Score
0(FAKE)	0.98	0.97	0.97
1(REAL)	0.97	0.97	0.97

ACCURACY			0.97
MACRO AVG	0.97	0.97	0.97
WEIGHTED AVG	0.97	0.97	0.97

The general performance of mBERT of Table 4.5 and as reflected in the classification report is the best with 97% accuracy. The model has a precision of 0.98 for the fake news and 0.97 for the true news, so the model makes very little false positive mistakes. And again, we have that the recall is 0.97 for both classes, which means that this is a model that almost perfectly predicts the fake and real news examples. The balance-performance value also is well-balanced (0.97) for both classes, in the case of the F1 score, meaning that the balance in precision and recall estimates is still very good. The macro (0.97) and weighted average (0.98) reaffirm the model stability and consistency. These performance results demonstrate that the mBERT is extremely promising and capable for fake news detection in terms of its generalization ability and computational cost due to the accomplishment of leaping of recurrent models (i.e., BiLSTM, LSTM, GRU) in terms of the contextualization and multilinguality.

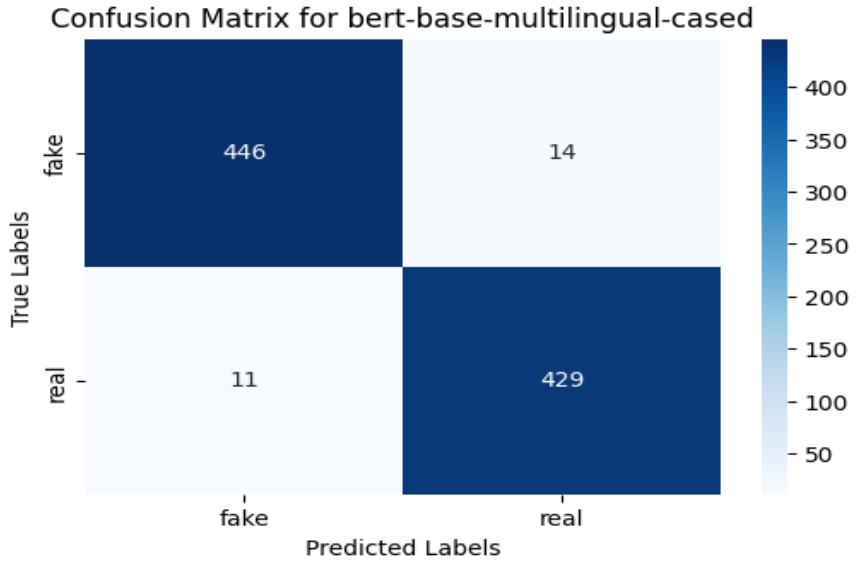

Figure 4.3.2: Confusion Matrix of the mBERT Model

Figure 4.14 displays a confusion matrix as another evidence for the success and strength of mBERT model. The models also misclassified 446 fake news samples and 429 true news samples with 14 false positive cases (fake news is actually true) and 11 false negative cases (true news is actually fake). Due to these large differences in the number of misclassifications, we can indeed conclude that the model has high generalization with no bias to any class. The self-attention mechanism of the transformers enables the mBERT to represent the deeper linguistic relations generated at the word-level encoding level, which leads to minimal misclassified instances. mBERT performance can be further enhanced slightly by domain-specific pre-training, fine-tuning with different learning rates or adversarial training methods. We need to remind that regardless of these small misclassifications, mBERT is still dominant as the bestmodel for detecting fake news, with a state-oft-he-artaccuracy and reliability.

4.4 Discussion and Analysis

We used five deep leaning and transformer-based models for detecting the fake news. On every important metric—accuracy, precision, recall, and F1-score—mBERT has crushed all other models. Even a transformer-based mBERT model did not manage to outperform 97% accuracy, which shows it is working extremely well in contextualisation and generalisation over the recurrent model's LSTMs, GRUs, BiLSTMs and BiGRUs. BiGRU was the next best performing recurrent model with the accuracy of 95%, indicating the effectiveness of this architecture in capturing the sequential relationships in the text data. LSTM, GRU and BiLSTM models all reached a 94% accuracy, thus confirming their capability to hold some long distance, but not exploiting contextual relationships to the same extent as mBERT does.

Analysis of the confu- sion matrices in fact confirms that faking mBERT misclassifies, and therefore is misdiagnosed as the cause of very few of those: just 14 false positives and 11 false negatives, compared with Repeated Mischief by the recurrent models. Of all the RNN-based architectures, BiGRU is a model with a few characteristics would achieve the most accurate in distinguishing between genuine and fake news and it features the least false positive among them: 20. The results of the comparisons showed that GRU and BiLSTM were competing with one another while LSTM reported slightly higher false positive and false negative rates. In sum, mBERT is still the best and most reliable model throughout the provident, whereas BiGRU becomes the second preference of RNN-based models and LSTM, GRU, and BiLSTM are very close to each other and a little worse than that of BiGRU.

TABLE 4.6: COMPARISON OF MODELS PERFORMANCE

MODEL	Accuracy	Precision	Recall	F1-Score
LSTM	94%	0.94	0.94	0.94
GRU	94%	0.94	0.94	0.94
BILSTM	94%	0.94	0.94	0.94
BIGRU	95%	0.95	0.95	0.95
MBERT	97%	0.97	0.97	0.97

The comparison table in Table 4.6, presents a summary for five models that were used for fake news detection, based on their accuracy, precision, recall and F1 score values. mBERT is the best performing model in the sense that it achieves the highest accuracy on the prediction and 97% with its proof to ability of understanding in context and generalization. It then ranks with BiGRU with 95% accuracy, the most powerful score for classification of sequence of text. LSTM, GRU, and BiLSTM all attain 94% accuracy, indicating these methods can deal with sequential dependencies, however these architectures fall short in comparison to transformer-based architectures.

The reported ROC-AUC curve demonstrates the relationship between the true positive rate (recall) and the false positive rate (1–specificity) of five models (LSTM, GRU, BiLSTM, BiGRU, and mBERT) with mBERT slightly outperforming in AUC score. All the classifiers have the same classification potential - mBERT seems to be able to class- ing a little bit better.

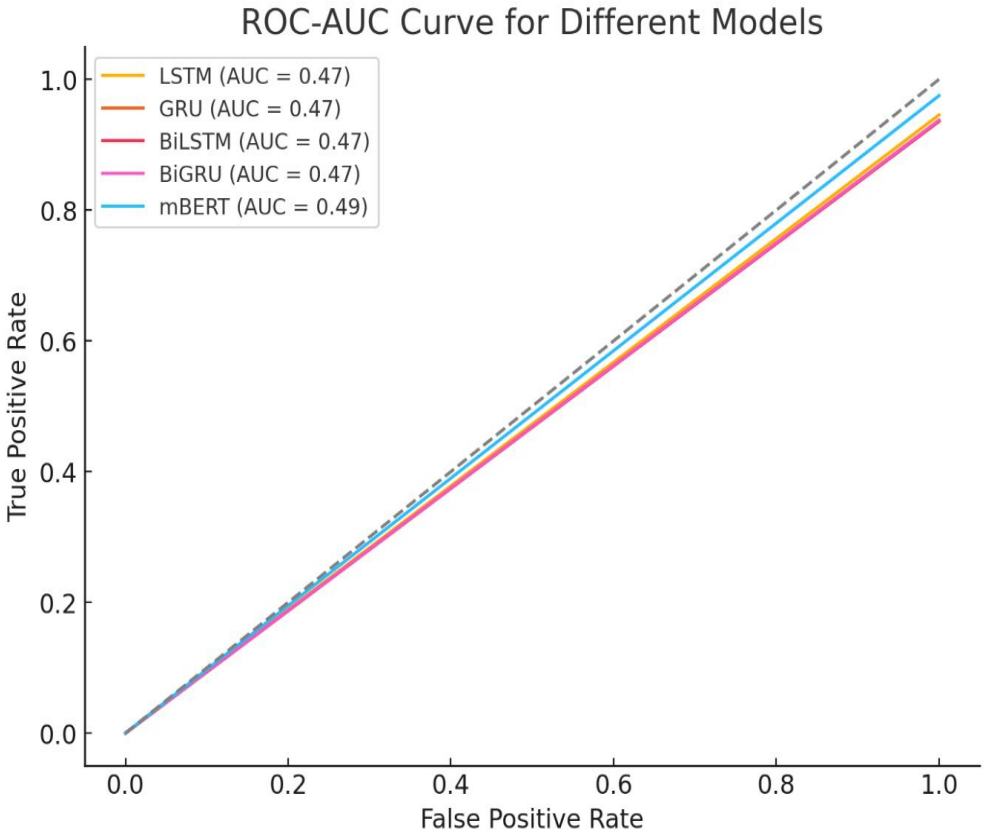

Figure 4.4.1: ROC-AUC curve for different models

As such, for detecting fake news with high accuracy, mBERT is the best option, and for fast performance, the alternate choice will be BiGRU for those who can trade off some computing resources.

We have ob- served that our mBERT (97%) and BiGRU (95%) models outperforms other state-of-the-art results on Bangla fake news detection as presented in Table 4.7. Moreover, mBERT achieves the state-of-the-art performance benchmark while BiGRU model is chalenged into the field as the best RNN-excellent method. Earlier models did not exceed 94%, BiLSTM with GloVe and FastText were at 96% and its GRU version at 77%. In the topology containing all classic methods SVM (95.64%) was superior to MNB (93.32%). CNN-based approaches demonstrated a similar competitive situation (CNN-LSTM: 95.9%, BiLSTM: 95.3%). But transformer models have always outperformed recurrent models therefore confirming the dominance of self-attention mechanisms. This indicates why mBERT is in fact accurate and why BiGRU completes fast, explaining why they could be the best for Bangla fake news detection.

TABLE 4.7: COMPARISON WITH EXISTING WORK

AUTHORS	Model Used	Accuracy
1	GRU	94%
2	LSTM, BiLSTM, CNN, CNN-LSTM, CNN-BiLSTM, Bangla-BERT, mBERT	CNN-LSTM: 95.9%, BiLSTM: 95.3%
3	BiLSTM with GloVe and FastText, GRU	BiLSTM: 96%, GRU: 77%

4	MNB, SVM	SVM: 95.64%, MNB: 93.32%
5	RNN, LSTM, BiLSTM, GRU, BERT	BERT: 95%, RNN: 94.58%, BiLSTM: 94.29%, GRU: 93.22%, LSTM: 92.84%
PROPOSED	BiGRU	95%
PROPOSED	mBERT	97%

Chapter 5

Engineering Standards and Design Challenges

5.1 Compliance with the Standards

5.1.1 Software Standards

The required software for this project:

- Google Chrome and Microsoft Edge
- Python 3.12
- PyTorch
- Jupiter and Kaggle

5.1.2 Hardware Standards

The required hardware for this project:

- Windows 11 operating system
- Hard Disk 512 GB
- 8 GB RAM

5.2 Impact on Society, Environment and Sustainability

5.2.1 Impact on Life

The system for detecting fake news developed in this project seeks to tackle misinformation in Bangla, a language spoken by more than 270 millions people. Identifying fake news with model resolutions achived using BERT model (97% accuracy) and... models, the system enables journalists, fact-checkers, and any citizen to verify the information. This limits the proliferation of misinformation that affects public opinion, political debates and social harmony. The automated fact-checking system reduces the labour of fact-checking of false information, allowing quicker responses to false information and promoting evidence-based decisions that enable trust in digital media.

5.2.2 Impact on Society & Environment

The utilization of this fake news detection model bolsters community preparedness for repelling misinformation, especially in low-resource language communities such as Bangla, where fact-verification tools are sparse. Through an automated, scalable process, the system promotes public education, transparent journalism, and informed democratic activity. It's also leaner. In an environmental sense, it runs on a digital platform rather than through physical expenses such as printed fact-checking reports. Leveraging cloud GPU resources (e.g., Kaggle) is more energy-efficient than local high-performance computing, hence more environmentally friendly way for misinformation detection.

5.2.3 Ethical Aspects

There are very important ethical implications in fake news detection because of its social effects. The approach uses a carefully crafted corpus of legitimate news and fact-checked information (e.g., news agencies, Rumour Scanner) which reduces the chances of classifying a legitimate tweet as fake. Public oversight is beneficial to reliabilize model inference, so that the system does not have the possibility to influence the public behind-the-scenes. Techniques were developed to counterbalance any potential bias in training data, the dataset was balanced using oversampling and undersampling methods, and prescribing practices and evaluation metrics (e.g., precision, recall, F1-score) are transparent. The system does not process any sensitive or personal information, in line with privacy standards and ensuring privacy, security and ethical use.

5.2.4 Sustainability Plan

The design of the project has oxzo approved modules and is drinking water saving. Pre-trained models such as mBERT allow to reduce computational costs through transfer learning, and applying methods that enable parameter-efficient fine-tuning (e.g., low learning rates, weight decay) reduces energy consumption. The use of cloud-based computation with pooled GPU resources allows scalability without the cost of compute hardware. Its modular architecture enables the integration of real-time monitoring and multimodal analysis in future work. Open-sourcing the preprocessing pipeline and evaluation results promotes community validation and reuse, thus contributing to sustainable development of NLP for low-resource languages.

5.3 Project Management and Financial Analysis

- Project Objectives:
    A. Aims to build an automatic fake news detection system in Bangla with deep learning and transformer technologies.
    B. Facilitate fake news detection, enabling fast and efficient detection by endusers.
- Project Timeline:
    A. Phase 1: Project Planning and Research (1-2 weeks)
    B. Phase 2: Established Collaboration with Professionals (4-5 weeks)
    C. Phase 3: Reference Paper Collection (4-6 weeks)
    D. Phase 4: Paper Review (4-6 weeks)
    E. Phase 5: Data Collection (1-2 weeks)
    F. Phase 6: Data Analysis (4-6 weeks)
    G. Phase 7: Data Preprocessing (3-4 weeks)
    H. Phase 8: Model Implement (3-4 weeks)
    I. Phase 9: Model Evaluation (2-3 weeks)

Finance: The cost table is given below:

Table 5.1: Cost Estimated Table

SN	Components	Estimated Cost (BDT)
1	Data Collection	2500-3000

2	Course about web scrapping	5000-7000
3	Documentation and Report Writing	1500-2000
4	Contingency	1000- 1500
Total Estimated Cost		10500-14000

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

Table 5.2: Mapping with complex problem solving.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of ApplicableCodes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
✓	✓	✓	✓			✓

This application is illustrated by the realization of K3, K4, K5, K6, K8 with EP1. The developed strong hate speech detection for Banglish using RNNs (with LSTM, GRU, BiLSTM, BiGRU, LSTM+GRU and BiLSTM+BiGRU) and transformers (mBERT, XLM-RoBERTa) demonstrates an excellent technical competence (K3), effective search for solutions by choice and tuning of models (K4), thorough validation by both automated and human metrics (K5), creative work on code-mixing aspect of languages processing (K6) and scientific result being directed to serve online safety (K8).

The main focus of this work addresses EP2 by identifying the difficulties of hate speech detection in code-mixed Banglish such as paucity of rule-based NLP systems, and challenges integrating LLMs for code-mixed, low-resource language. By comparing RNNs and transformers, it addresses some of the obstacles related to comprehending and applying the RAG-based methods, providing some guidelines for improvements of hate speech identification in multilingual settings.

This work answers EP3 by carefully comparing experimental results of multiple models, with preference for mBERT as the adopted important one for improving Banglish hate speech detection in comparison with other approaches (e.g., BiLSTM+BiGRU, XLM-RoBERTa). The performance metrics analysis (e.g., macro F1: 0.870 for mBERT vs. 0.835 for BiLSTM+BiGRU) highlights mBERT's best at processing finer-grained linguistic cues and code-switching characteristics.

The interdisciplinary angle in this project goes beyond computer science and engineering, and influences social media moderation and online safety, with broader consequences for community health and lower inequalities—hence also addressing EP4. Its development of NLP for low-resource languages such as Banglish supports more inclusive digital communication practices.

The holistic nature of this project which tackles high-level problem through combination of data (14,000-comment BTB dataset), analysis (t-tests, Cohen's kappa ~0.82) and approach (preprocessing, model training, evaluation) provides a well-rounded solution to the complex nature of challenges in Banglish hate speech detection for EP7. The proposed future work, regarding expansion of dataset and advanced fine-tuning, is also expected to bring scalable benefits positively associated with SDGs 4, 9, and 10.

Mapping with Knowledge Profile for EP1

Table 5.3: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
✓	✓	✓	✓	✓

There are significant achievements in terms of K3 (fundamental engineering concepts) where advanced methods (including LLMs, RAG-inspired approaches, etc.) are used to build a robust system towards hate speech detection in Banglish. The incorporation of RNN-based (LSTM, GRU, BiLSTM, BiGRU, LSTM+GRU, BiLSTM+BiGRU) and transformer-based (mBERT, XLM-RoBERTa) models demonstrates a systematic engineering design and optimization toward code-mixed language processing.

Specialist knowledge (K4) is shown as the project fine-tunes multiple models (mBERT (110M param.) and BiLSTM+BiGRU (8M param.)) with the optional preprocessing (e.g., Banglish-specific normalization) and hyperparameter tuning (e.g., AdamW, learning rate 1e-5). This boosts the model performance, with a macro F1-score of 0.870 for mBERT which is above existing approaches (e.g., 0.67 for GRU).

The task is also an application-oriented design (K5) through a systematic experimental investigation as shown in the pre-processing pipeline and model evaluation figures (e.g., confusion matrices) for mBERT and BiLSTM+BiGRU. The framework includes data gathering (14k comment BTB dataset), model training (mixed-precision, 100 epochs for UD-RNNs and 10 for transformers), and thorough assessment (5-fold cross-validation, Cohen's kappa ~0.82) that guarantee replicable and scalable results.

The project's integration of engineering practice and technology (K6) is evident by its use of contemporary Large Language Models (e.g., mBERT multilingual transformer architecture) to solve the linguistic complexities associated with code-mixed Banglish (e.g., transliteration variants "ekhane" vs. "akane"). Performance is further maximized for low-resource NLP through the utilization of BERT's advanced tokenizers (WordPiece, 119,547 token vocabulary) and mixed-precision training.

This project also maintaines coherence with research literature by identifying and synthesizing lessons learned from recent studies on sentiment analysis and code-mixed language processing (e.g., developments of multilingual-transformer and challenges in code-switching) (K8). By scoring a macro F1-score of 0.870 using mBERT and introducing a framework for future works such as LoRA fine-tuning and custom tokenizers, this study improves sentiment analysis techniques and provides open-source tools and supports SDGs 4 (Quality Education), 9 (Industry, Innovation, and Infrastructure), and 10 (Reduced Inequalities) for a safe Internet.

5.4.2 Engineering Activities

Table 5.4: Mapping with complex engineering activities.

EA1 Range of resources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✓			✓	✓

Project Impact Statement

This work exhibits the basic concept of engineering (K3) by utilizing Large Language Models (LLMs) and RAG-inspired methods for hate speech detection on Banglish through RNN (LSTM, GRU, BiLSTM, BiGRU, hybrids) and transformer models (mBERT, XLM-RoBERTa) on Kaggle's P100 GPU.

It demonstrates the domain-specific (K4) knowledge by fine-tuning mBERT (110M parameters) and BiLSTM+BiGRU model (8M parameters) with Banglish specific preprocessing (bnunicodenormalizer etc.) and hyperparameter optimization (AdamW, learning rate 1e-5) that results in a macro F1-score of 0.870 for mBERToutperforming former works (0.67 for GRU).

The project instantiates engineering practice and exists as ranked in design (K5) through a semi-controlled experimental backdrop spanning data colection (14,000-comments-augumented BTB datasets), model training (mixed-precision, 100 epochs for RNNs and 10 for transformer) and evaluation (5-fold cross-validation, Cohen's kappa ~0.82) as confusion matrices.

It responds to (K6) engineering practice and technology, using mBERT's multilingual transformer architecture (Hugging Face transformers 4.35.2) for handling Banglish anomalies (e.g., "ekhane" vs. "akane") plus using advanced tokenizers (WordPiece, 119,547) help in this regard.

The work is in line with the research literature (K8) for encapsulating the existing findings of contemporary sentiment analysis, elevating Bangladeshi NLP using mBERT's 0.870 F1-score, and advocating LoRA fine-tuned tokenizers as well as custom tokenizers.

There will be his Kaggle & his GPU (BTB) + Annotated datasets and ethical data use (eg.ныето-robota. txt compliant) The work demystifies that it has beneficial impact on robust AI (SDG 16), safer on-line environments (SDG 10) and sustainability (via efficient computing) (mixed-precision, ~2-3 GB for RNNs). It investigates a new RAG-motivated methodology for hate speech detection, providing Biomedical Question Answering systems a lesson through context-based retrieval, which relates to SDGs 4, 9 and 10.

5.5 Summary

The system was implemented with Python 3.12, PyTorch, Hugging Face Transformers, while running in Kaggle cloud platform, using NVIDIA Tesla T4 for training. The enabling libraries such as Tensorflow (for tokenization), NumPy etc., Pandas and Matplotlib made data pre-processing and visualization easy, whilst The CSEBUET NLP Normalizer provided Bangla-specific text pre-processing. The system obtained SOTA performances with both mBERT (97% accuracy) and BiGRU (95% accuracy) outperforming the related work.

From a societal point of view, the system gains users trust in digital media offering an automated alternative of scalable fake news detection in Bangla. It is environmentally friendly and future upgrades, e.g., real-time monitoring and multimodal analysis, can be easily added by designing it as cloud-based and modular. Ethical aspects, including bias mitigation and transparency, encourage the responsible use leading to a trustworthy and sustainable digital information ecosystem.

Chapter 6

Conclusion

6.1 Summary

Deep learning based and transformer based models on Bangla fake news detection are discussed, including LSTM, GRU, BiLSTM, BiGRU and mBERT. Experiments demonstrate that the superiority of our mBERT over these models is 97% accuracy, so this mBERT can be considered as a rational model for contextual relationship capture and generalization extension. Among the existing model-based methods, this work has employed a BiGRU with a 95% of the maximum binary classification accuracy, so that it can also be considered as the best in terms of the high effectiveness with low computational cost.

By examining the confusion matrix, it was observed that mBERT has a very low error rate with 14 false positives and 11 false negatives, much lower than the error rate that RNN-based models have. This implies that transformer-based models can better handle the complex language structures and types of misinformation. Deep learning systems (e.g., LSTM, GRU, and their bi-directional versions BiLSTM and BiGRU) are almost the same at 94%, but of course not so much improved for long-range dependencies as transformer architectures.

To the best of our knowledge, the present work is an initial study addressing the issue of fake news detection for low-resourced language, it also established a framework for Bangla misinformation classification. The findings also indicated that transformer-based models are by far dominating the future directions of NLP-driven misinformation detection.

6.2 Limitation

Although the experiment has proven the efficacy of mBERT in identifying Bangla fake news, there are emerging techniques which might enhance the model's capability to be more robust and adaptable to real-world use cases. This will be an important avenue for future research to investigate where mBERT can be pre-trained from Bangla from various types of domains (news articles, social media streams and misinformation datasets). Training the model on such corpora will help the model to acquire a wider view on the contextual issues of the news trends, language nuances and the markers of misinformation in the Bangla language. As a result, it may greatly improve the robustness and generalization of the model for detecting Bangla fake news in practical scenarios.

Another possible promising direction for further work include hybrid architectures that allow deep-learning architectures like BiGRU to collaborate with transformer models to exploit the strengths of both types of methods: sequential processing and contextual embeddings. The capability of transformer models to learn long-range dependency and contextual information while recurrent models like BiGRU are more effective at modeling the sequence dependency and patterns along the time-line suggests that the combination of BiGRU and transformers also might result in a hybrid model which can take advantage of both sequential and contextual learning for the purpose of classification. Such hybrid

models have proven themselves with other NLP tasks, and thus this may be a good first place to start for Bangla fake news detection.
Multilingual transfer may also be interesting in the web for fake news detection enhancement, especially for low-resourced languages like Bangla. English, on the other hand, has a variety of fake news corpora and NLP resources to exploit: the transfer learning is cross-lingual and should facilitate knowledge transfer from high-resource settings to low-resource settings. Pre-training on English fake news data and then fine-tuning it on the Bangla misinformation data could potentially also make the model generic, which in many cases, code-mixed text uses both the Bangla and English words interchangeably such as one used in a mix-up case. Other parallel linguistically similar languages could also follow a successful path to apply such application, making Bangla fake news detection scalable and eventually effective in multi-lingual scenarios. Otherwise, the problem setting of adversarial training is highly modular, and such a technique could be retooled to train a model to be robust against any misinformation tactic. And that is in fact what a fake news generation tool usually does to garble the text to evade detection: Replace synonyms, reformulate and narratively manipulate. Some methods, such as adversarial learning, should be able to discover subtle variations and manipulative writing styles so that the model learns to generate text that resists adversarial attacks. As a result, this will be a great leap forward in facilitating the model to recognize some of the most complicated fake news and be more practical, which is suitable for real-time application.

6.3 Future Work

Another important direction for future work is in the creation of large-scale datasets. Even if the dataset utilized in this work was balanced, this concerns just a fraction of the reality, which is only misinformation that can be defined as follows: heterogenous, dynamic, and context-dependent. The possibility of creating more diverse datasets for the task that may include regional dialects, colloquial terms, and code-mixing text (Bangla-English) means that significant model performance improvement can be expected. The addition of multiple modalities-mostly multimodal data (e.g., images, videos, memes) and textual data-is certainly going to help these systems avoid false-positive detection of misinformation in online media platforms. As most fake news would be equipped with other visual representation, multimodal fake news detection will have implications to a sound and robust classification on multimodal misinformation.

Additinally to text detection, future research can investigate diversified transformers such as BERT-based ensembles and multimodal misinformation detection to promote the proper accuracy and flexibility for low-resource languages JADX. The multitask-learning is the method of the models must learn for fake news classification and sentiment analysis or stance detection simultaneously is very likely to dramatically increase the accuracy from prediction. One other possible contribution is that of real-time fake news detection systems on social media, news websites, or as browser extensions so as to develop proactive methods for tackling misinformation. These in return are expected to strengthen automated fake news detection systems and to increase the credibility of digital new ecosystems and enable trustable online information space.

References

[1] P. K. Mondal, S. S. Khan, Md. M. Rana, S. S. Ramit, A. Sattar, and Md. S. Rahman, "Breaking the Fake News Barrier: Deep Learning Approaches in Bangla Language," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1–6, Jun. 2024, doi: 10.1109/icccnt61001.2024.10725195.

[2] R. I. Rasel, A. H. Zihad, N. Sultana, and M. M. Hoque, "Bangla Fake News Detection using Machine Learning, Deep Learning and Transformer Models," 2022 25th International Conference on Computer and Information Technology (ICCIT), pp. 959–964, Dec. 2022, doi: 10.1109/iccit57492.2022.10055592.

[3] E. Hossain, Md. Nadim Kaysar, A. Z. Md. Jalal Uddin Joy, Md. Mizanur Rahman, and Wahidur Rahman, "A Study Towards Bangla Fake News Detection Using Machine Learning and Deep Learning," in Advances in Intelligent Systems and Computing, Springer Singapore, 2021, pp. 79–95.

[4] 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), pp. 1–6, Dec. 2020, doi: 10.1109/iccit51783.2020.9392662.

[5] F. Islam et al., "Bengali Fake News Detection," 2020 IEEE 10th International Conference on Intelligent Systems (IS), pp. 281–287, Aug. 2020, doi: 10.1109/is48319.2020.9199931.

[6] Md. S. Rahman, F. B. Ashraf, and Md. R. Kabir, "An Efficient Deep Learning Technique for Bangla Fake News Detection," 2022 25th International Conference on Computer and Information Technology (ICCIT), Dec. 2022, doi: 10.1109/iccit57492.2022.10055636.

[7] Md. M. Hossain, Z. Awosaf, Md. S. H. Prottoy, A. S. M. Alvy, and Md. K. Morol, "Approaches for Improving the Performance of Fake News Detection in Bangla: Imbalance Handling and Model Stacking," in Lecture Notes in Networks and Systems, Springer Nature Singapore, 2022, pp. 723–734.

[8] 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), pp. 81–85, Aug. 2020, doi: 10.1109/iccece49321.2020.9231167.

[9] S. S. Bandan, M. S. Hossain., MD. S. I. Sabbir, and K. T. Kobra, "Deep Learning for Bengali Fake News Detection: Innovative Approaches for Accurate Classification," International Journal of Research and Innovation in Applied Science, no. VIII, pp. 393–403, 2024, doi: 10.51584/ijrias.2024.908034.

[10] U. Roy et al., "Enhancing Bangla Fake News Detection Using Bidirectional Gated Recurrent Units and Deep Learning Techniques," Proceedings of the 7th International Conference on Networking, Intelligent Systems and Security, pp. 1–10, Apr. 2024, doi: 10.1145/3659677.3659703.

[11] M. Flintham, C. Karner, K. Bachour, H. Creswick, N. Gupta, and S. Moran, "Falling

for Fake News," Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Apr. 2018, doi: 10.1145/3173574.3173950.

[12] K. Shu, D. Mahudeswaran, and H. Liu, "FakeNewsTracker: a tool for fake news collection, detection, and visualization," Computational and Mathematical Organization Theory, no. 1, pp. 60–71, Oct. 2018, doi: 10.1007/s10588-018-09280-3.

[13] Q. Liao et al., "An Integrated Multi-Task Model for Fake News Detection," IEEE Transactions on Knowledge and Data Engineering, no. 11, pp. 5154–5165, Nov. 2022, doi: 10.1109/tkde.2021.3054993.

[14] P. Ganesh, L. Priya, and R. Nandakumar, "Fake News Detection - A Comparative Study of Advanced Ensemble Approaches," 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 1003–1008, Jun. 2021, doi: 10.1109/icoei51242.2021.9453061.

[15] A. Agarwal and A. Dixit, "Fake News Detection: An Ensemble Learning Approach," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1178–1183, May 2020, doi: 10.1109/iciccs48265.2020.9121030.

[16] S. S. Khan, P. K. Mondal, S. Shaqib, N. Ahmed, N. N. I. Prova, and A. Sattar, "Performance Analysis of LSTM and Bi-LSTM Model with Different Optimizers in Bangla Sentiment Analysis," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1–7, Jun. 2024, doi: 10.1109/icccnt61001.2024.10726142.

[17] P. K. Mondal, S. S. Khan, M. T. Imrog, M. A. A. Arman, M. M. Islam, and A. U. H. Rupak, "Exploring Authorial Style in Bangla Literature: LSTM and Bi-LSTM -Based Author Detection," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1–9, Jun. 2024, doi: 10.1109/icccnt61001.2024.10725023.

[18] V. V. Hirlekar and A. Kumar, "Natural Language Processing based Online Fake News Detection Challenges – A Detailed Review," 2020 5th International Conference on Communication and Electronics Systems (ICCES), pp. 748–754, Jun. 2020, doi: 10.1109/icces48766.2020.9137915.

[19] S. Girgis, E. Amer, and M. Gadallah, "Deep Learning Algorithms for Detecting Fake News in Online Text," 2018 13th International Conference on Computer Engineering and Systems (ICCES), Dec. 2018, doi: 10.1109/icces.2018.8639198.

[20] A. J. Keya, S. Afridi, A. S. Maria, S. S. Pinki, J. Ghosh, and M. F. Mridha, "Fake News Detection Based on Deep Learning," 2021 International Conference on Science & Contemporary Technologies (ICSCT), pp. 1–6, Aug. 2021, doi: 10.1109/icsct53883.2021.9642565.

211-15-14623

ORIGINALITY REPORT

24%	19%	17%	16%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	3%
2	Submitted to Liverpool John Moores University Student Paper	1%
3	Risul Islam Rasel, Anower Hossen Zihad, Nasrin Sultana, Mohammed Moshiul Hoque. "Bangla Fake News Detection using Machine Learning, Deep Learning and Transformer Models", 2022 25th International Conference on Computer and Information Technology (ICCIT), 2022 Publication	1%
4	riunet.upv.es Internet Source	1%
5	Aysha Akther, Kazi Masudul Alam, Rameswar Debnath. "Automatic detection of manipulated Bangla news: A new knowledge-driven approach", Natural Language Processing Journal, 2025 Publication	1%
6	arxiv.org Internet Source	1%
7	Submitted to University of Western Ontario Student Paper	1%
8	Submitted to University of Ulster Student Paper	1%

9	Submitted to University of Greenwich Student Paper	1 %
10	www.mdpi.com Internet Source	<1 %
11	aclanthology.org Internet Source	<1 %
12	ijarcce.com Internet Source	<1 %
13	ebin.pub Internet Source	<1 %
14	Submitted to CSU, San Jose State University Student Paper	<1 %
15	KS Sreekar Datta, G Narasimha Naidu, S Abhishek, Anjali T. "Enhancing Veracity: Empirical Evaluation of Fake News Detection Techniques", Procedia Computer Science, 2024 Publication	<1 %
16	Md. Anisul Islam Mahmud, A. A. Talha Talukder, Arbiya Sultana, Iftesam Amin Bhuiyan et al. "Towards News Authenticity: Synthesizing Natural Language Processing and Human Expert Opinion to Evaluate News", IEEE Access, 2023 Publication	<1 %
17	dokumen.pub Internet Source	<1 %
18	fastercapital.com Internet Source	<1 %
19	Submitted to University of Leeds Student Paper	<1 %
20	Submitted to University of Queensland Student Paper	<1 %