Assessing the Effectiveness of Modern Deep Learning Architectures in Multi-Label News Categorization

Ali Haider, Asma Junaid
School of CS & IT, Institute of Management Sciences, Peshawar, Pakistan
*Correspondence: asmajunaid0109@gmail.com

Citation | Haider. A, Junaid. A, "Assessing the Effectiveness of Modern Deep Learning Architectures in Multi-Label News Categorization", IJIST, Vol. 8, Issue. 2 pp 546-566, April 2026

Received | March 03, 2026 **Revised |** April 05, 2026 **Accepted |** April 09, 2026 **Published |** April 12, 2026.

The rapid spread of fake news online poses a serious threat to public trust, making automated detection systems increasingly necessary. The proliferation of AI-generated content has intensified this challenge, with the number of AI-enabled fake news sites increased tenfold in 2023 alone [1]. This study is the first to compare seven deep learning architectures — CNN, LSTM, Bi-LSTM, RNN, Bi-RNN, GRU, and Bi-GRU — under identical experimental conditions on the same dataset, providing a reliable and unbiased performance benchmark. All models were trained on a publicly available dataset of 44,919 labeled news articles (23,502 fake, 21,417 real) using the Adam optimizer (learning rate = 0.001), binary cross-entropy loss, a batch size of 64, and early stopping, with sequences padded to 500 tokens and an 80/20 train-test split. Each model was evaluated using accuracy, precision, recall, F1-score, and misclassification count. GRU achieved the best overall performance, with 99.91% accuracy, precision, recall, and F1-score, recording only 8 misclassifications on 8,980 test samples. CNN ranked second with 99.82% accuracy and 16 misclassifications, followed by Bi-LSTM at 99.74% with 21 misclassifications. The LSTM achieved 99.57%, while the RNN reached 99.21% Bi-RNN performed the weakest at 96.80% accuracy with 287 misclassifications, the highest error count among all models. Notably, the standard GRU outperformed its bidirectional counterpart, Bi-GRU (99.69%), suggesting that increased architectural complexity does not always yield better performance. These findings demonstrate that GRU is a strong, lightweight choice for real-time fake news detection and offer practical guidance for selecting deep learning models in misinformation filtering systems.

Keywords: Deep Neural Networks, Fake News Detection, Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) Architecture.

Introduction:

In the current digital era, where the dissemination of information is rapid through websites, social media, and messaging applications, fake news has become a global issue. Many people now prefer to get their daily news from online platforms, but not all the news available online is true. Researchers often use the term "fake news" alongside related terms such as misinformation, disinformation, hoax, and rumor, which refer to different forms of false information dissemination. Misinformation refers to the sharing of false content without considering intent, often caused by poor fact-checking or mistakes such as wrong labeling of images. Disinformation refers to the false information that is purposely made and shared to mislead or manipulate, often linked to propaganda. Rumors and hoaxes describe false stories that are deliberately made to look real, even though the facts are inaccurate or fabricated [2]. The purpose of fake news is to mislead, sway, or create confusion among people. Important sectors like politics, public safety, and even individual opinions might get influence by this. False claims about vaccines resulted in widespread opposition during the COVID-19 pandemic, which hindered immunization process [3]. *It is now increasingly difficult for individuals to distinguish* the difference between the fake and real news because of AI-generated deepfakes and bots. Fake news has the ability to quickly shape people's opinions and decisions making this situation particularly concerning. Fake news may negatively impact societies and continue to sway opinions if it left unchecked.

Researchers and practitioners have investigated this issue from a variety of perspectives and tried to address fake news through human moderation, fact-checking websites, and machine learning, but all of these approaches are insufficient to keep up with the vast amount of information that is shared online every second around the world. So, they started looking for automated ways that can help in classifying misinformation. deep learning architectures have achieved outstanding results in the detection of fake news [4]. Deep Learning architectures such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTMs), and Recurrent Neural Networks (RNNs) have been used to analyze the text, source, headline, and metadata of news articles and these architectures performed better as compared to other traditional methods [5]. Furthermore, deep learning architectures are trained in such a way that they can easily learn the underlying trends and patterns of fake news, including bias, inconsistency, and sensationalism. [6]. In order to make predictions, these models do not need humans input instead they learn patterns from large datasets and by integrating such models into online platforms, we can minimize the spread of fake news [7]. Recent studies further confirm the severity of this problem. According to NewsGuard, the number of AI-enabled fake news sites increased tenfold in 2023, and approximately half a million deepfake videos were shared on social media in that same year [1]. [8] note that the rapid advancement of social media platforms has facilitated the spread of misinformation at an unprecedented rate, making it difficult for individuals to distinguish fact from fiction. Similarly, [9] highlight that the scale and speed of online misinformation have made manual fact-checking and rule-based filtering fundamentally insufficient for the volume of content produced daily.

In spite of these developments, we are still confused about how different deep learning architectures distinguish from other architectures when used to detect fake news because each architecture has their own pros and cons when handling textual data and to understand which model perform best in different situations and why. Additionally, some previous studies used only a single architecture to identify news as real or fake or they might does not test these models on different datasets. This study solves this problem by comparing various deep learning architectures alongside one another instead of using a single architecture. It takes a close examination on the way each model is performing and what type of patterns they are picking up and how good or bad they are performing.

The literature has a definite research gap: most published works compare only a single or two architectures separately under varying experimental conditions, and it is hard to compare models reliably [10]. It is uncommon to compare bidirectional variants with their unidirectional counterparts when applied to the same set of settings, and crucial implementation details including tokenizer parameters, batch size and hardware configurations are often not included, which constrains reproducibility [10]. This paper directly fills these gaps by enacting and comparing seven architectures, using a completely standardized and documented experimental protocol.

The primary goal of this study is to implement various deep learning architecture on the Fake and Real news dataset and subsequently compare their outcome to understand how each model functions. We are using news datasets that are publicly available on Kaggle for the training purposes. After model's training, we will assess their performance through evaluation measures such as accuracy, recall, F1-score, precision and number of misclassifications and identify the most effective model that can provide reliable results. We will compare each architecture's performance and their ability to deal with various news types in order to understand its strengths and weaknesses. We intend to develop and test these deep learning models, as they can quickly and accurately detect misinformation, in order to produce a system that assists people in avoiding misleading information. Researchers, online platforms, and regular users may find this helpful. Additionally, we want to ensure that the models we build are reliable and are capable of performing well in real-life situations. The sole subject matter of this research is not only about the use of technology but making the internet safer, by educating people more about the difference between the fake and real news. Through this paper, we want to stop the spread of misinformation by utilizing better and smarter technologies.

Research Objectives:

This study will be guided by the following specific and measurable objectives:

To use seven deep learning models (CNN, LSTM, Bi-LSTM, RNN, Bi-RNN, GRU, Bi-GRU) to classify binary fake news on a standardized set of 44,919 labeled articles.

To compare every model on five quantitative measures: accuracy in the test, precision, recall, F1-score, and the overall number of misclassifications, which is further broken down into false positives and false negatives.

To determine which architecture has the best classification accuracy and the minimum number of prediction errors given the same training conditions.

To understand whether bidirectional processing is always effective in boosting the performance compared to unidirectional counterparts of all gating architectures.

The structure of this research paper is as follows: Section 2, introduces the previous studies regarding the methods of detecting fake news with the assistance of machine learning and deep learning algorithms. Section 3 addresses the proposed methodology, including dataset, data preprocessing methods and deep learning architectures, which were used, such as CNN, RNN, Bi-RNN, LSTM, Bi-LSTM, GRU and Bi-GRU. Section 4, shows the outcomes and discusses the performance of these models. Lastly, Section 5, concludes the study by presenting our findings.

Research Contributions:

This section clearly points out the difference and the extension of the current study to the existing research on the topic of detecting fake news using deep learning.

Multi architecture benchmarking: This work implements and compares seven different models using the same dataset, preprocessing pipeline, and hyperparameter settings, and evaluation metrics, unlike most previous studies that only test a single or two architectures [11][12]. This makes sure that any differences in performance seen are due to true architectural attributes and not experimental variability.

Detailed metric reporting: This paper will also report on accuracy but will also include the precision, recall, F1-score, and the number of misclassifications by false positive and false negative per model. This detail is not provided in a number of similar studies [12][13] and also gives the practitioners the opportunity to choose models on the basis of particular error tolerance needs.

New empirical evidence on bidirectionality: An important conclusion of this research is that bidirectional extensions do not consistently outperform their unidirectional counterparts. Bi-GRU (99.69%) underperforms standard GRU (99.91%), and Bi-RNN (96.80%) is the weakest model overall. This directly counters one popular supposition of the literature and offers practical advice in regard to architecture choice [14].

GRU as a lightweight high-accuracy solution: This paper shows that GRU has 99.91% accuracy, which is higher than CNN, LSTM, and all the variants of bidirectional network, and it has fewer trainable parameters than LSTM. This makes GRU a good candidate in resource efficient real- time misinformation detector systems [14][15].

Fully reproducible experimental design: Complete implementation details — including tokenizer parameters, padding strategy, batch size, learning rate, optimizer, early stopping configuration, and hardware specifications — are reported to enable replication and extension of this work [10].

Literature Review:

This section reviews earlier research on the detection of fake news, using deep learning and machine learning techniques. To compare performance on different datasets, researchers have tested a variety of models, including Naive Bayes, SVM, Random Forest, CNN, LSTM, GRU, and their bidirectional. In order to create an accurate and reliable detection systems, it is necessary to compare the effectiveness of these techniques and to find which model is most effective for detecting fake news.

[11] conducted research to optimize deep learning models through hyperparameter tuning in order to improve their fake news detection system. This research focuses on enhancing the performance and reliability of system that can correctly identify misinformation on online platforms. They used optimized CNN and LSTM models on two different datasets with political news article labeled as fake or real to achieve research goals. To increase model accuracy, they used HyperOpt for hyperparameter tuning. The optimized CNN was able to extract significant features of the text and attain an accuracy of 93.1 percent, and the LSTM detected patterns of sequences and attained 89.96. The performance was evidently increased when compared to the non-optimized models. Also, the paper has highlighted the importance of tuning to enhance detection and use of clean and labelled data. Their research supports using deep neural networks to combat false information on the internet._Nevertheless, this paper only considers two architectures and excludes GRU and any bidirectional variations. The accuracy figures reported are significantly lower than those obtained in the current study; this could be due to differences in size of data set or tuning depth. There is also no misclassification breakdowns reported in the study, which restrict the analysis of errors.

[16] focused on utilizing ensemble deep learning models to detect fake news in the Arabic language and by dealing with the unique challenges, their work makes it much easier to identify misinformation. They used the Bi-LSTM and Bi-GRU architecture together and therefore developed a powerful hybrid model that enabled the ability to comprehend the text structure and meaning easily. This hybrid model was tested using two Arabic datasets, AFND and ARABICFAKETWEETS which demonstrated high accuracy. They also transformed words into useful vectors to be trained with FastText. They had a better hybrid model than other models such as CNN-LSTM and RNN-CNN. They also tried other models of transformer such as BERT and XLNet but the Bi-LSTM-Bi- GRU model worked better. This experiment demonstrated that combining various deep learning models can go a long way in enhancing fake

news detection and offers a strong base to develop future systems that are multilingual and can be adapted to various contexts. While notable for its multilingual scope, this study fuses Bi-LSTM and Bi-GRU into a single hybrid model, making it impossible to isolate the individual contribution of each component. The present study addresses this gap by evaluating each architecture independently under identical conditions.

[13] worked on utilizing a deep neural networks approach based on LSTM to identify fake news and investigates how deep learning and machine learning techniques can help in the more effective and efficient detection. It considers different techniques for identifying false information on the internet, especially on platforms like Facebook, Instagram were misinformation spreads quickly. The main focus of this study is on sequential models which has the ability to recognize hidden patterns in the news content which cannot be detected by traditional machine learning models, and it highlights the importance of working with advanced models to boost the performance. They found that deep learning handled the text flow much better than other classifiers like Random Forest and Logistic Regression. Before training, they ensured that the dataset was balanced to prevent biassed outcomes and used TF-IDF to convert text into numerical representation after cleaning it by eliminating extra words and then breaking text into tokens. Moreover, they mentioned that the next generation models will contain pictures or videos since fake news doesn't necessarily involve only texts. This paper is restricted to one LSTM model and it does not discuss GRU, Bi-LSTM and bidirectional versions. The 94% accuracy on the Kaggle dataset reported, although useful as a baseline, is also far lower than the 99.91% of GRU in the current study, which validates the importance of the larger comparison performed here.

Using deep learning and machine learning architectures, [17] investigated COVID-19 fake news detection and public sentiment analysis. Their approach identified fake pandemic-related content with a 97% accuracy. The authors draw attention to how quickly false information spread on social media during the pandemic and point out the urgent need for accurate and effective models to prevent the spread of false information. They evaluated deep learning models like CNN, LSTM, Bi-LSTM, and GRU as well as models like Naive Bayes, SVM, and Random Forest. According to their results, deep learning models particularly CNN and Bi-LSTM performed better at identifying fake news, understanding text patterns and processed more quickly and accurately as compared to traditional methods. To improve the efficacy of such systems, the study proposed that future research should use social media features, multi-language data, and real-time fake news detection tools. Bangyal et al. is among the most directly comparable studies to this work, as it evaluates multiple architectures including GRU. However, it does not include Bi-GRU or Bi-RNN, leaving the cross-variant comparison incomplete. Their finding that CNN and Bi-LSTM outperformed GRU contrasts with our results; this difference may be attributable to the domain-specific nature of the COVID-19 dataset used.

[2] introduced an advanced deep learning approach to address the growing problem of fake news spreading quickly across social media platforms. In a bid to enhance the automatic process of detecting fake news- Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), the authors suggested a hybrid model. This hybrid model was also experimented with two datasets, that is, ISOT and FAKES and was built in such a way that it could be in a position to identify significant pattern of words and context in the article, it was found to be more effective than CNN, RNN, or any other machine learning model, which was seen to work individually. The study tested the extrapolation of the model on several datasets and could conclude that the model was only useful in case the data were known, and the model was pathetic in case the data were unknown. The authors have outlined limitations of what was done in other research such as the inability to select features effectively and lack of flexibility and have indicated that the models that are capable of addressing the different news types must

be developed. According to the outcome of their investigation, deep learning is applicable to finding fake news, but further studies are needed to make it more consistent across various data sets.

[18] studied the way deep learning models could be applied to detect fake news using as the sole information the text of an article. They tested three models (LMST, CNN and fine-tuned BERT) to identify how they can categorize the news articles as fake or true ones. The models were focused on the analysis of the title of the articles and primary content but not on the social or user-related metadata. They had an experiment that BERT was the best in every circumstance as it was able to generate an accuracy of over 93%. They were trained and tested on TI-CNN and Fake News Corpus. The authors used Bayesian optimization of the hyperparameters of the LSTM and CNN models though BERT was fine-tuned through the best practices. Overall, the study has underscored that deep learning models, in most cases, transformer-based models can be used to detect fake news more effectively with greater accuracy using text only and with scaling. The following work cycle was to gather the new data so as to increase the model performance and competence in handling the invisible data.

To counterfeit the widespread spread of false information on the Internet, [19] proposed a deep learning model that discovers fake news. To employ an ensemble deep learning model, which involves fusing the superiority of a large number of neural models and thus improving the accuracy of the detection framework of the fake news, the authors used the LIAR dataset. Through their experiment, they prove that detection performance can be significantly improved using the textual and attribute-based features. Instead, they applied dense neural network to other metadata features and Bi-LSTM-GRU on the attribute of textual statement. The research entailed data pre-processing consisting of the initial phase of addressing missing data, normalization, and natural language processing (NLP) including tokenization and lemmatization. The authors discovered that the addition of other features to the model did not affect the performance of the model and highest accuracy of 89.8% was achieved in the case of the statement attribute alone. Their model was far better compared to the methods used in the previous research on the same data. The authors jumped to the conclusion that despite the fact that their model was highly successful, the improved generalization is possible with the assistance of further research using the assistance of various data.

[12] proposed that the process of verifying the facts and the traditional approach have drawbacks; thus, the deep learning tool should be applied to enhance the process of fake news detection. The authors were able to come up with a detection system that was able to distinguish the real and fake news based on the Long Short-Term Memory (LSTM) deep learning model and compare the performance of LSTM to the other machine learning models i.e., the Support Vector Machine (SVM) and the K-Nearest Neighbors (KNN). The LSTM model was more accurate with 88 percent and recalled higher with 90 and high precision of 86. It is observed by the authors that LSTM was quite effective because it could understand the sequence of words in news articles. They demanded a viable system that can recognize fake news in the right and effective way. Moreover, the authors suggested that this model was supposed to be tested on larger data sets and that as many features as possible should be added to improve the performance of this system. The significance of their move is in mollify bogus news on the net.

[20] identified the need to have improved tools capable of looking at the bigger picture of the news. A hybrid deep learning architecture that is based on sequential learning strategies alongside content-based approaches is suggested to address this as the authors suggested. They tried to enhance the process of detecting fake news using feature extraction, which included N-gram and TF-IDF, and text pattern learning, i.e., LSTM and BERT. Their model was geared towards improving the representation of content involving news in a way that would be easy to distinguish between real or fake contents. This combination method was also experimented on this The WELFAKE and Kaggle Fake News data set and the results were 96.8% accuracy on

WELFAKE and 94% accuracy on Kaggle Fake News because deep learning and traditional feature extraction were jointly applied such that results were improved over the earlier models. Future research, according to Kausar et al., should concentrate on testing using a variety of datasets, investigating multimodal features, and enhancing the model's transparency for improved decision-making.

[21] worked on using deep neural networks to better identify fake news in English and Turkish text. They build a transformer-based model that integrates news content and social media opinions and can quickly detect fake news. Their model, that take motivation from the BART architecture, combines social context elements like user comments, engagement metrics, and credibility with news elements like headlines and body text. Additionally, they introduced a weak supervision method that automatically assigns labels in order to address the problem of limited labelled data. The accuracy of the model was 74.8% and it outperformed many deep learning and traditional baselines particularly in the early detection of the characteristics. The main advantage of the current research is that it will be able to spot the misinformation a few minutes after its release and change it to the constantly-changing method with the fake news. Their work underscores the importance of including the social context and content to the process of detecting fake news. The authors proposed to consider adding more social specificities and multilingual skills, improved multi-labeling detection and exploration of novel data in the future studies.

[15] evaluated CNN, LSTM, CNN-LSTM, and CNN-GRU for fake news detection on the ISOT and FakeNewsNet datasets. Their CNN-GRU model achieved 99.97% accuracy on FakeNewsNet, demonstrating the power of combining convolutional and recurrent layers. However, their study does not include standalone GRU or bidirectional variants, and the two-dataset setup differs from our single standardized benchmark. [22] similarly proposed a hybrid GRU-CNN architecture for news text classification, reporting improved accuracy over standalone CNN and GRU baselines. While confirming the complementary strengths of these architectures, their study does not isolate GRU performance under controlled conditions as done in the present work.

However, even though the learning algorithms in detecting misleading news based on deep neural network, machine learning and a combination of the two have made significant progress, several limitations do exist. Experimenting with the LSTM and BERT, [20] obtained high results, but, they did not consider other models and ignored them. Similar to [2], which employed CNN and RNN, [19] employed a BiLSTM-GRU model without further comparing the models which does not permit them to understand comparative weaknesses or strengths of the model in a better way. BERT based models were giving plausible results as they were used by [18] however there was also the issue of the model generalizability. Similarly, [12][11] used LSTM and optimized CNN/LSTM models to improve the performance but their studies did not include any comparison at large scales. While some researchers, like [21][16], widened the scope of work to multilingual detection, most of studies were still based on isolated or limited methods. Collectively, such gaps attract attention to the fact that there has not been a single framework that would enable the simultaneous testing of multiple deep learning models in equal conditions. To deal with this, our research uses and compares various architectures in the same framework based on evaluation measures such as the number of misclassifications, accuracy, precision, recall, and F1score. This provides a better perspective of the comparative strengths and weaknesses of different models and a better perspective on how they respond when compared to each other.

Methodology:

Techniques using machine learning and deep learning were utilized in several recently published studies in order to detect misinformation. Even though these techniques have yielded positive outcomes but their performance depends on the dataset, news source and type of the model applied. Whereas certain models are more effective in cases where they are trained with large datasets, others are more effective with lower volumes of data. To ensure that fake news spreads in various forms, it is essential to test and compare various deep learning models under the same conditions. To determine which model is the most efficient to detect fake news, the research uses different deep learning models and then takes a closer look at their performance. Our training and evaluation process was structured to train and evaluate various deep learning architecture types used to detect real and fake news. The initial step was the preprocessing of the data whereby it involved tokenizing the data into numerical sequences and cleaning the text and removal of noise in the text. These data were purged to training (80%) and testing (20%). The entire experimental pipeline follows a series of five steps: (1) Dataset Preparation - raw CSVs are loaded, labeled (Fake = 0, Real = 1), and randomized with a fixed random seed; (2) Text Preprocessing - text cleaning, tokenization, and sequence padding are performed on all samples in the same way; (3) Model Construction — seven architectures are built in TensorFlow/Keras, each sharing a common embedding layer and sigmoid output; (4) Training and Validation — all models are trained under identical hyperparameter settings with early stopping, using an internal 20% validation split of the training set; (5) Evaluation — all models are assessed on the same held-out test set using five performance metrics.

The authors have utilized different deep learning structures including CNN, LSTM, RNN, GRU and their two-way counterparts. In order to provide fair comparison of each of the models, all of them were trained in the same environment, after which each of the models was evaluated using evaluation metrics such as accuracy, precision, recall, F1-score and the misclassification count. The measures help in identifying each model's advantages and disadvantages. All the findings were compared finally towards the realization of the most effective and dependable model of fake news identification. The method can be replicated and expanded in future researches. The proposed research technique is presented in figure 1.

Dataset:

We uses the publicly available Fake and Real News Dataset from [23], which comprises of Fake.csv file and True.csv file. The first file, Fake.csv, has 23,502 fake news articles, while the second file, True.csv, has 21,417 true news articles. The true news dataset is made up of verified, actual news reports, whereas the fake news dataset most likely includes facts which are fabricated or false. The articles are labeled (Fake=0, Real=1) which shows whether the news is authentic or fake. Figure 2 illustrates the class distribution, the difference in counts happened because the total number of fake samples in the dataset is slightly less than what was originally mentioned. To ensure that the model doesn't learn based on any particular order in the dataset, the data is shuffled after labeling for the sake of randomness. The difference in the number of articles between the two files is small, ensuring a balanced dataset for analysis. Such datasets are typically utilized by researchers and developers to enhance machine learning models that are usually trained with such files so as to differentiate between genuine and fraudulent news. Key features of Fake and Real News Dataset is shown in Table 2.

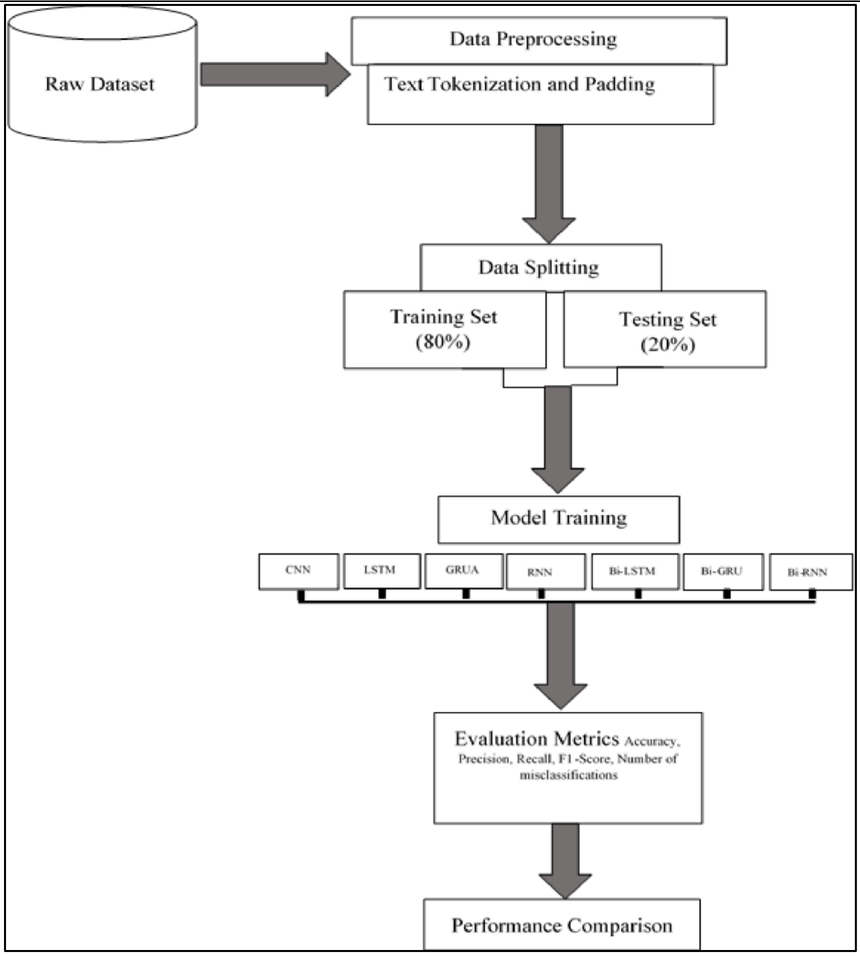

Figure 1. Proposed Research Methodology

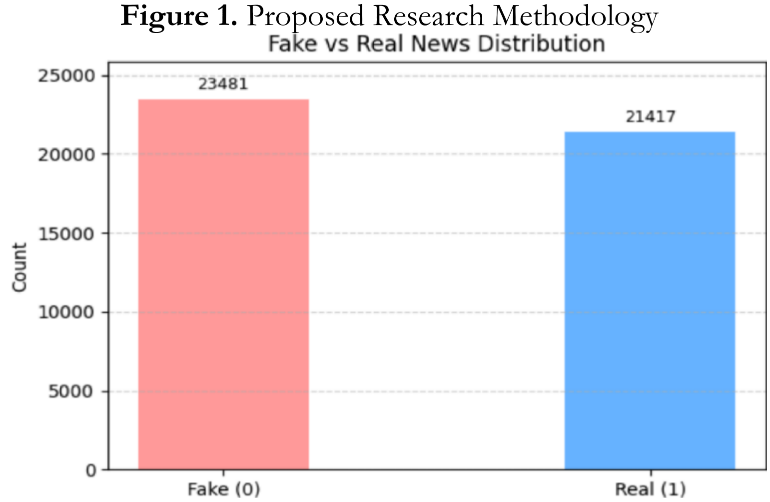

Figure 2. Class Distribution: Fake vs Real News

Data Preprocessing:

Before being fed into any deep learning model, the fake and real news data underwent several preprocessing steps that were performed to prepare the data to be analyzed. First, all text was converted to lowercase to ensure uniformity. Regular expressions are used to remove unnecessary content like text in square brackets, URLs and HTML tags. The non-alphabetic characters were removed too and additional spaces were substituted by one space to ensure that the text remains in order. Words that do not add much significance to the analysis process (like the, and, a) were filtered out to give attention to the meaningful words. After cleaning the text, it is then tokenized with Keras Tokenizer with a vocabulary size of the 5,000 most common words, which boosts the performance of the model and alleviates noise. Padding is also employed to ensure that all sequences are exactly 500 tokens in length since models need input to be the same length. Longer texts are truncated and shorter texts are padded with zeros to ensure that the model is able to deal with input easily in uniform. Finally, the dataset is split into training and testing sets with 8,980 samples assigned to testing and 35,918 samples assigned to training.

Training data shape: (35,918, 500)
Testing data shape: (8,980, 500)

Table 1 below sums up the preprocessing steps. The Keras Tokenizer was set up with a vocabulary size of 5,000 words and an out-of-vocabulary token to handle unseen words. Each sequence was then padded or truncated to a constant length of 500 tokens (zeros were used as padding), resulting in each input being the same shape. We have used a fixed random seed (42) to randomize the dataset before dividing the data to prevent any ordering effects. During training, 20% of the training data was kept aside as a validation set to track model performance and decide when to stop training early.

Table 1. Summarized preprocessing steps

Input	Combined news text (title + body text)
Step 1	Convert all text to lowercase
Step 2	Remove URLs, HTML tags, and content within square brackets (regex patterns)
Step 3	Remove non-alphabetic characters; replace multiple spaces with single space
Step 4	Remove English stopwords (NLTK stopword list, ~180 words)
Step 5	Tokenize using Keras Tokenizer: vocab_size = 5,000; oov_token = '<OOV>'
Step 6	Convert cleaned tokens to integer sequences
Step 7	Pad / truncate all sequences to length = 500 (post-padding, post-truncation; pad value = 0)
Step 8	Label encoding: Fake = 0, Real = 1
Step 9	Shuffle dataset with fixed random seed (random_state = 42)
Step 10	Split: 80% training (35,918 samples) / 20% test (8,980 samples)
Output	Integer sequences of shape (N, 500) with binary labels {0, 1}

Table 2. Key Features of the Fake and Real News Dataset

Column	Description	Example
title	Headline of news article	"Trump Just Closed the Office That Coordinates Russia Sanctions"
text	body text of news article	"With Donald Trump late on implementing the new sanctions..."
subject	Topic of news article (e.g., news, politics)	"News"
date	Date of publication	"October 27, 2017"

Deep Learning Model Architectures:

We developed deep neural network models to meet the objectives of our research. Their working is to convert the text to vectors (Embedding Layer) and it is terminated with a dense layer that will tell whether a news is real or fake. All architectures are trained with Adam optimizer, binary cross-entropy loss and concept of early stopping (stop after 3 rounds without improvement). Such designs are compatible with typical standards in the study of fake news detection (e.g., LSTM, GRU, CNN). The common feature of all models is their input structure: an Embedding layer (input dimension = 5,000; output dimension = 128; input length = 500; trained end-to-end) and architecture-specific layers, with a final Dense output layer with sigmoid activation as a binary classifier.

Below is a brief explanation of each model:

Convolutional Neural Network (CNN):

Convolutional neural networks (CNNs) are useful to identify text patterns, including particular phrases or groups of words that commonly occur in fake news. In this work, text was initially converted to word embeddings, and then filtered through Conv1D filters to scan important n-grams. Max-pooling dimensionality-reduces and prevents overfitting. The actual prediction was done by the last layers. However, CNN is not able to recall sentence structure or extensive context, and it is excellent at quick detection and works well on short text. Recent studies with CNN+LSTM hybrids indicate improved pattern recognition in fake-news problems. Hyperparameter-tuned optimized CNNs have obtained significant accuracy improvements [11]. The CNN model employed in this study has an embedding layer and a convolutional layer with ReLU activation, a max pooling layer, a flatten layer, and several dense layers and a sigmoid activated output to perform binary classification. The proposed CNN based text classification model is detailed and the structure is shown in Figure 3.In this study, the CNN uses 128 filters of kernel size 5 with ReLU activation, followed by global max pooling, a Dense layer (64 units, ReLU), Dropout (rate = 0.5), and sigmoid output. CNN inference is fully parallelizable, making it faster than recurrent models at test time — a practical advantage for high-throughput applications [15].

Long Short-Term Memory (LSTM):

LSTM models are constructed to deal with longer text in which previous portions of a sentence influence subsequent meanings. In this project the LSTM retain long term context with the help of internal memory cells and gating. This assisted in comprehending the connection of the words throughout the sentence. The LSTM required additional time to train, although it performed better on full-length articles. In a 2023 study, LSTM was estimated to be 94 percent accurate on the identical Kaggle fake-news dataset [13]. LSTM layer, as Figure 4 demonstrates, is effective in its ability to deal with the sequential nature of the text, which is why it can be used in the task of fake news classification. In our research, the LSTM model consists of one LSTM layer (128 units, return sequences = False) and one Dense layer (64 units, ReLU), Dropout (0.5) and a sigmoid. Three gating mechanisms of LSTM input, forget and output gates enable selective retention of information that is contextually relevant, which is advantageous with long news articles.

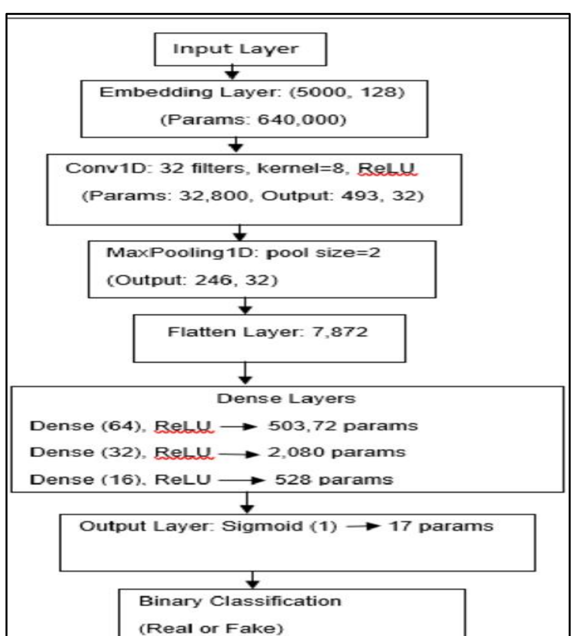

Figure 3. Architecture of the Proposed CNN Based Text Classification Model

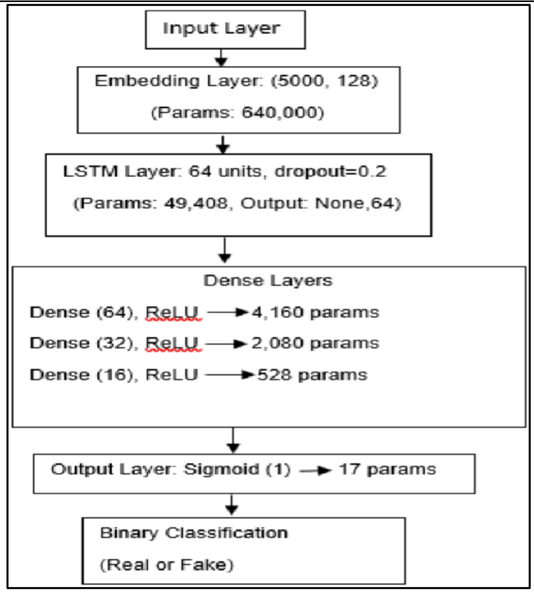

Figure 4. Architecture of the Proposed LSTM Based Text Classification Model

Bidirectional LSTM (Bi-LSTM):

Bi-LSTM reads the text in two directions both forward and backward, as shown in Figure 5, helping it in understanding what came before and after each word. The results of both the forward and backward LSTM paths were combined and passed on to dense layers in our research. It did take more training time but gave improved accuracy. The attention-based Bi-LSTM model shows remarkable performance, outperforming other architectures in terms of accuracy (97.66%) and other key metrics [8]. In our study, Bi-LSTM achieved higher accuracy, recall and precision than LSTM.

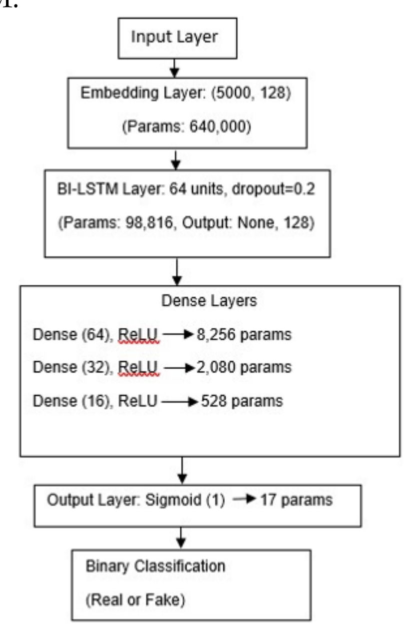

Figure 5. Architecture of the Proposed Bi-LSTM Based Text Classification Model

Simple Recurrent Neural Network (RNN):

The current project used a simple RNN to determine the performance of more simplistic models. The simple RNN processes information sequentially, storing past information. The study adopted an embedding layer, followed by an RNN layer and dense layers for output, as demonstrated in Figure 6. Although very efficient, RNNs tend to have problems with longer text since they have forgotten the previous sections of the text, which is referred to as vanishing gradient problem and short long-term memory. The recurrent neural networks (RNNs) architecture has difficulties in acquiring long-term memories, because of vanishing and exploding gradients [24]. In our case, RNN performed moderately well on short inputs but gave limited results on longer text. Despite this, it is useful for quick baseline evaluations and environments with limited resources due to its computational simplicity and minimal training time.

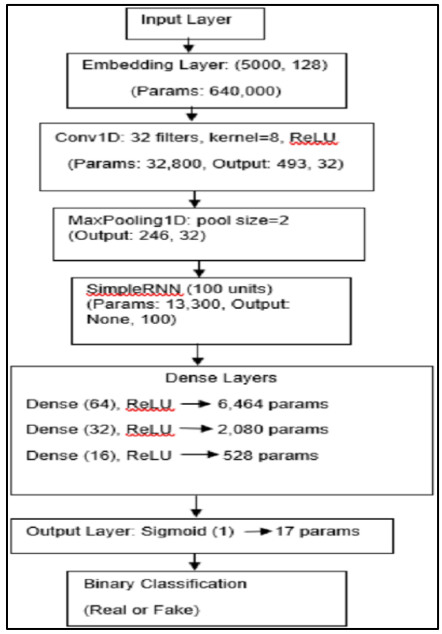

Figure 6. Architecture of the Proposed RNN Based Text Classification Model

Bidirectional Recurrent Neural Network (Bi-RNN):

BiRNNs have two simple layers of RNN: The first reads the text backward and forward and the second reads it forward and backward. This dual processing aids the model by considering both in the previous and later context in the interpretation of meaning of each word more comprehensively and then fuses the outputs and comes up with a final prediction. It can be applied when it comes to text classification where, perhaps because of the context in both ends of a sentence, the results are better than without the understanding of it. BiRNNs can still be utilized to maintain the clarity of the information flow in the course of training and deployment. Figure 7 below shows the proposed architecture of the Bi-RNN based model of text classification.

Gated Recurrent Unit (GRU) & Bidirectional GRU (Bi-GRU):

Gates Recurrent Units (GRUs) is a gate-based decision-maker on what to remember or what to forget, and it was faster and faster, which provided it with the same performance as LSTM, but with less parameters. GRU provided similar results as LSTM in our work but it was able to learn faster. GRU model architecture is represented in figure 8, whereby the gates are used to recall valuable text information. The Bi-GRU model, shown in Figure 9, reads text in both directions and provided the best balance between speed and accuracy. Bi-GRU in our experiments was almost as accurate as Bi-LSTM and requires less training time, which supports the argument that Bi-GRU can be as efficient and effective as LSTM-based models [25]. GRU was chosen as the main subject of this paper due to its good trade-off between accuracy and the cost of computation. GRU has two gates, reset and update, instead of three in comparison to LSTM, which decreases the number of trainable parameters but does not affect sequence modeling abilities [14]. This causes GRU to be more appropriate in resource-constrained environments like mobile applications and edge servers. The empirical finding that GRU (99.91) is expected to perform better than Bi-GRU (99.69) is also in line with the results that the extra model complexity of the bidirectionality does not always result in a better performance, as it was observed in [26].

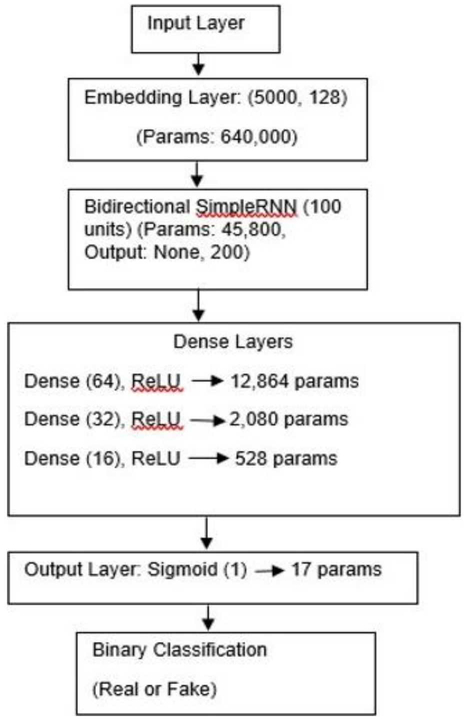

Figure 7. Architecture of the Proposed Bi-RNN Based Text Classification Model

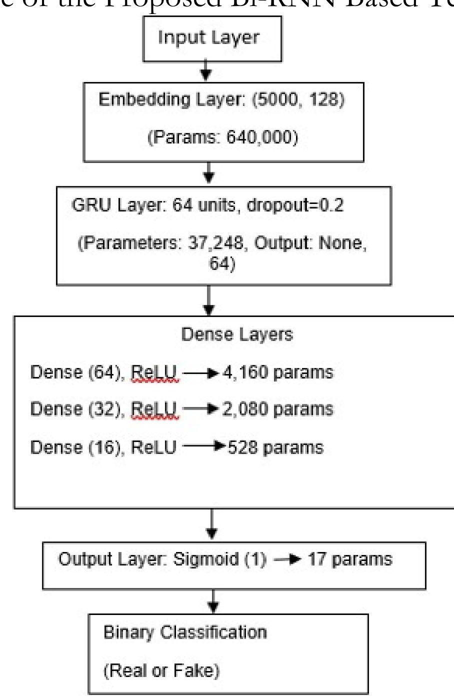

Figure 8. Architecture of the Proposed GRU Based Text Classification Model

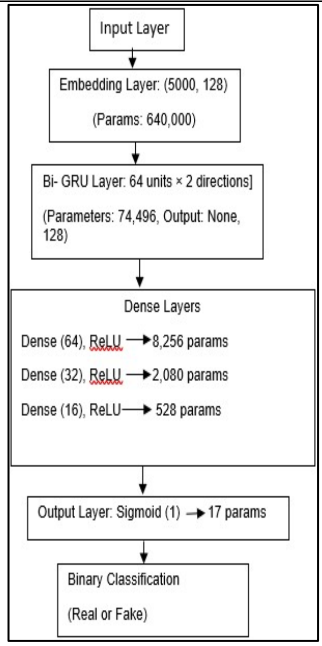

Figure 9. Architecture of the Proposed Bi-GRU Based Text Classification Model

Implementation Details:

All seven models were trained under the following identical configuration to ensure a fair comparison:

Optimizer: Adam with learning rate = 0.001

Loss function: Binary cross-entropy

Batch size: 64

Maximum epochs: 10 (with early stopping)

Early stopping: Patience = 3 epochs, monitoring validation loss

Embedding layer: input dimension = 5,000; output dimension = 128; input length = 500; trained end-to-end

Framework: TensorFlow 2.x / Keras

Hardware: NVIDIA CUDA-enabled GPU, 16 GB RAM, Intel Core i7 processor

The models were also trained in a single deterministic run with a specific random seed (42). This is because the early stopping and the use of fixed random seed provides reproducibility to all the reported results. Keras was trained on a 20% validation split of the training data (7,184 samples), which was sampled randomly out of the 35,918 training samples, and used to monitor convergence during the training. This should be done to many runs with various seeds in the future to report mean accuracy and standard deviation [10].

Results and Discussion:

This section presents the comparative evaluation of various deep learning models implemented for binary classification of news as "real" or "fake". The models were evaluated using key performance metrics: test accuracy, precision, recall, F1-score, and the number of misclassifications (i.e., sum of false positives and false negatives). Evaluation metrics like accuracy, precision, recall are used to evaluate classification models for balanced datasets [27]. Accuracy tells us how many total predictions are correct out of all the predictions made.

$$Accuracy = \frac{T_p + T_n}{T_p + F_p + T_n + F_n}$$

Precision measures the number of positive cases that are correctly predicted.

$$Precision = \frac{T_p}{T_p + F_p}$$

Recall shows how well the model identifies actual positives cases.

$$Recall = \frac{T_p}{T_p + F_n}$$

The F1-score is most useful when data is imbalanced, and it is a harmonic average of precision and recall.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Table 3. Evaluation metrics of deep learning models on the test dataset

Model	Test Accuracy	Precision	Recall	F1 Score	FP + FN (Misclassifications)
CNN	99.82%	99.81%	99.81%	99.81%	16 (8+8)
GRU	99.91%	99.91%	99.91%	99.91%	8 (5+3)
Bi-LSTM	99.74%	99.58%	99.88%	99.73%	21 (16+5)
LSTM	99.57%	99.76%	99.32%	99.54%	44 (11+33)
Bi-GRU	99.69%	99.72%	99.62%	99.67%	27 (12+15)
RNN	99.21%	99.39%	98.94%	99.16%	71 (26+45)
Bi-RNN	96.80%	96.50%	96.75%	96.63%	287 (149+138)

As noted by [28], although some metrics like accuracy and F1-score are affected when one class significantly outweighs the other, in balanced datasets these same metrics provides more reliable and meaningful evaluations of both positive and negative predictions. Among all the deep learning models, the GRU model achieved the highest performance with a test accuracy of 99.91%, and precision, recall, and F1-score all recorded at 0.9991, as shown in Figure 10 and Figure 11. Figure 13 shows that, it made only 8 misclassifications, indicating that the model has learned the underlying patterns in the data, not just memorized the training examples and its ability to work well even when the input text is messy, incomplete, or has errors. The CNN model followed closely with a 99.82% accuracy, showing balanced precision and recall (both 0.9981), as shown in Figure 10 and Figure 11. As shown in Figure 12 and Figure 13, the model's 16 misclassifications were evenly divided, containing 8 false positives and 8 false negatives.

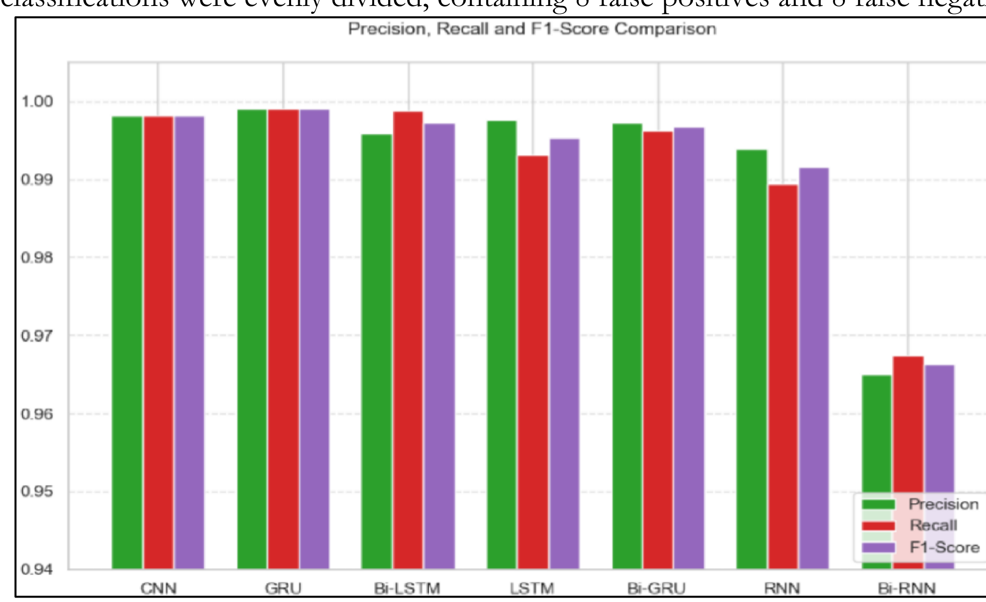

Figure 11. Comparison of different deep learning architectures based on Precision, Recall, and F1-Score

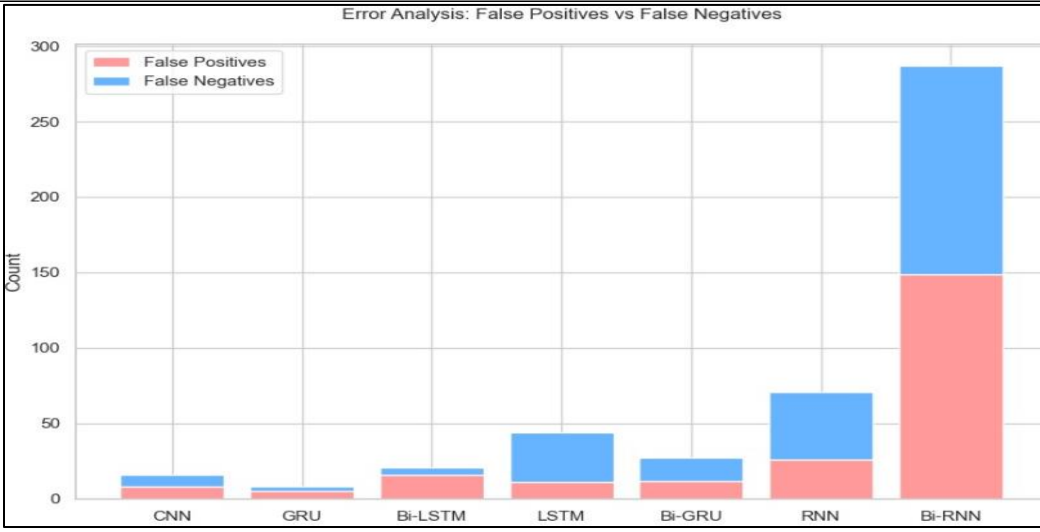

Figure 12. Error analysis of deep learning architectures showing the comparison of false positives and false negatives for each architecture

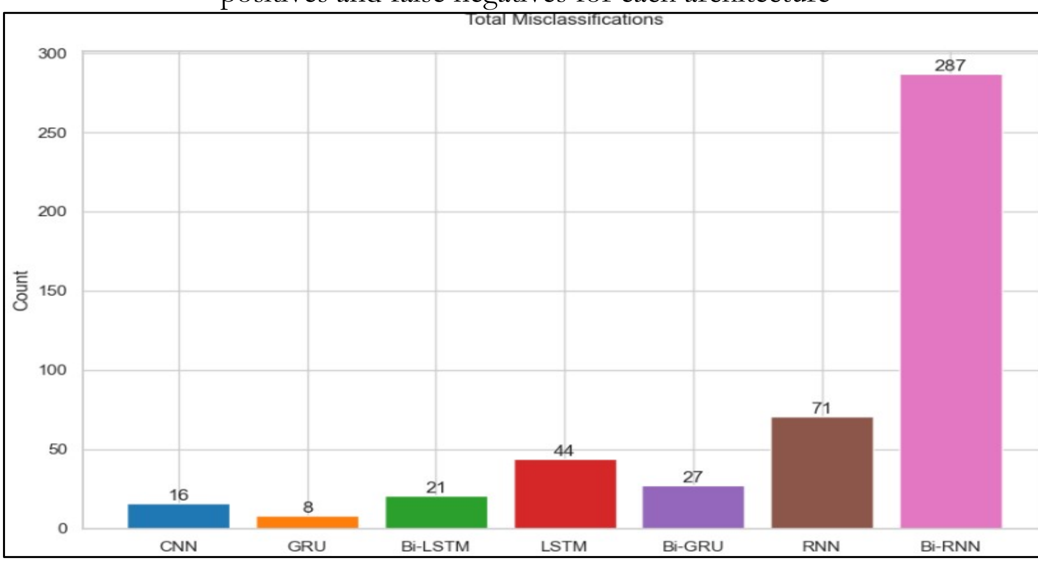

Figure 13. Total number of misclassifications for each model, highlighting overall prediction errors across different DL architectures

As illustrated in Figure 11, the Bi-LSTM model also performed well, obtaining a high F1-score of 0.9973 and it has the ability to process text in both forward and backward directions helped it comprehend the context better. The Bi-GRU model, however, did not function as well. Even so, its F1-score of 0.9967 produced decent outcomes, as illustrated in Figure 11, compared to the standard GRU, which had fewer misclassifications, its performance is slightly worse (see Figure 13). Although it is often believed that bidirectional models perform better because they can capture both past and future context, this is not always true. As stated in [26], the intricacy of the data and the trade-off between model depth and overfitting risk can affect how well Bi-GRU and related architectures perform. This means that simpler model like GRU can be more effective than models in which we have added bidirectionality. Interestingly, the LSTM model had made 44 misclassifications despite having a good accuracy of 99.57%, as illustrated in Figure 10 and Figure 13, which implies that it struggled a bit picking up long term patterns. With 99.21% accuracy and 71 misclassifications, RNN performed slightly worse, as illustrated in Figure 10 and Figure 13, demonstrating its inability to handle long sequences since it lacks memory gates like LSTM.

The Bi-RNN model showed the weakest performance, with only 96.80% accuracy and the highest number of misclassifications (287), as illustrated in Figure 10 and Figure 13. Simple RNN architecture was not able to detect complex patterns in fake news. These results align with previous studies that found vanilla RNNs have weaknesses that include vanishing gradients, which reduce their capability of recognizing long run dependencies in textual data [29]. Table 3 demonstrates that gated models such as GRU and LSTM, and their bidirectional versions, are overall superior to simple ones such as RNN. The model that was most effective in this study in terms of classification of fake news was the GRU based model. Since deep learning models are able to capture contextual relationships in text, they have been demonstrated to perform quite well in detecting fake news particularly when using memory units such as LSTMs and GRUs [30].

Regarding result confidence: Each model was trained in a single deterministic run with a fixed random seed. The consistent ranking of gated models (GRU, CNN, LSTM) above non-gated models (RNN, Bi-RNN) across all five metrics — and the large performance gap between GRU (99.91%) and Bi-RNN (96.80%) — provides strong qualitative confidence that the observed differences reflect genuine architectural effects rather than random variation. Future work should conduct multiple runs with different seeds to formally report standard deviation and confidence intervals.

Component Contribution Analysis:

Although a formal ablation study was not conducted - since each model is a discrete architectural choice rather than a modular component of a single system - the controlled comparison allows the contribution of each architectural element to be assessed conceptually.

Effect of gating mechanisms: Comparing RNN (99.21%) to LSTM (99.57%) and GRU (99.91%) isolates the contribution of gating. Adding gates consistently reduces misclassifications, confirming that the ability to selectively retain or discard sequence information is critical for long-text classification [14].

Effect of bidirectionality: Bi-GRU (99.69%) underperforms GRU (99.91%), and Bi-RNN (96.80%) is the weakest model overall. This suggests that bidirectionality introduces additional parameters that increase overfitting risk on this binary classification task without proportional accuracy gains, consistent with [26].

Convolutional effect: CNN (99.82) is better than LSTM, RNN, Bi-GRU, and Bi-RNN, and it is clear that for this dataset, local n-gram features that Conv1D filters capture are very discriminative without modelling long-range dependencies [15].

Practical Implications:

The results of the current study have a number of practical implications on organizations implementing real-world fake news detectors.

GRU for high-accuracy pipelines: With 99.91% accuracy and only 8 misclassifications on 8,980 samples (error rate: 0.09%), GRU is well-suited for automated content flagging where a small residual error can be handled through human review.

CNN on real-time high through putt systems: CNN is appealing to real-time systems due to its parallel nature of inference and the fact that it is 99.82 percent accurate, which would introduce bottlenecks when used sequentially in the recovery of repetition inference [15].

Avoiding Bi-RNN in production: The significantly lower accuracy of Bi-RNN (96.80%) and its high misclassification count (287) suggest it should not be used in production-level systems without substantial further development.

Lightweight deployment: GRU requires fewer parameters than LSTM, making it better suited for resource-constrained environments such as mobile applications and edge devices.

Conclusion and Future Work:

Conclusion:

This study was aimed to compare and evaluate the performance of various deep learning architectures to check the authenticity of news articles. We utilized a Fake and Real News Dataset on the Kaggle and trained several deep learning frameworks, which are CNN, LSTM, Bi-LSTM, RNN, Bi-RNN, GRU, and Bi-GRU. We also assessed the models after training and testing in terms of accuracy, precision, recall, F1-score and the misclassifications. Although we did not find any of the models worse than the other, we found out that some were more reliable and accurate than the others. To illustrate, GRU presented the best overall performance with the highest test accuracy and the least misclassifications. Conversely, such models as Bi-RNN possessed poorer accuracy and greater misclassifications. This fact is one of the main benefits of this study, as it will be possible to compare various deep learning models with each other on the same dataset and on the same evaluation methods. This facilitated our easier understanding of not only the models that do better but also their response to same input data.

To see how each model performs in detecting fake news, we used multiple models rather than just one or two. Some models gave more consistent results but others were more effective at detecting subtle patterns in the news. This will be able to guide others to the best model that can meet their needs. We would like to determine in the future how these models will perform across regions, topics and datasets such as in other languages and by combining ensemble methods and hybrid models to further enhance accuracy. Also, real-time fake and real news can be used to further promote upcoming research and enhance the accuracy of such systems.

Future Work:

Based on the result and limitation of this study, the following are the recommendations that can be made to future research:

Comparison based on transformers: BERT, RoBERTa, and DistilBERT, in turn, need to be compared within the same experiment to find out whether their performance improvement can justify such a significant increase in the cost of computation.

Cross-lingual evaluation: The models will be evaluated on non-English datasets (such as Arabic, Urdu, and Hindi) to test cross-lingual generalizability.

Hybrid architectures: GRU and CNN should be considered as ensembles due to the complementary properties of local feature detection (CNN) and sequential context modeling (GRU) that have been shown in this work [22].

Multimodal fake news detection: Adding metadata on images, video content, or social engagement indicators to text could enhance the ability to detect more recent types of fake news that use visual manipulation [9].

Multi-run statistical assessment: Future research needs to use multiple runs of the training with varied random seeds and to report average accuracy and standard deviation to describe statistically sound performance assertions.

Real-time implementation: Implementing the GRU and CNN models on social media streams in real-time would confirm the findings in production environments and expose the challenges that are not reflected in offline assessments.

Acknowledgment:

The authors extend their gratitude to the anonymous reviewers for their constructive feedback and suggestions, which helped us make our research better.

Author Contributions:

Ali Haider initially came up with and planned the experiments, did the experiments, did the computations, made figures and tables, wrote or reviewed rough drafts of the paper, and gave approval to final draft.

Asma Junaid did the experiments, did the computations, composed and reviewed rough drafts of the paper, and gave the final draft her approval.

Conflict of interests:

The authors state that there is no conflict of interest.

References:

[1] "AI and Misinformation | 2024 Dean's Report." Accessed: Apr. 21, 2026. [Online]. Available: https://2024.jou.ufl.edu/page/ai-and-misinformation

[2] Jamal Abdul Nasir, Osama Subhani Khan, "Fake news detection: A hybrid CNN-RNN based deep learning approach," *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 1, p. 100007, 2021, doi: https://doi.org/10.1016/j.jjimei.2020.100007.

[3] D. Allington, B. Duffy, S. Wessely, N. Dhavan, and J. Rubin, "Health-protective behaviour, social media usage and conspiracy belief during the Covid-19 public health emergency," *Psychol Med*, vol. 51, no. 10, pp. 1763–1769, Jul. 2020, doi: 10.1017/s003329172000224x.

[4] Linmei Hu, Siqi Wei, "Deep learning for fake news detection: A comprehensive survey," *AI Open*, vol. 3, pp. 133–155, 2022, doi: https://doi.org/10.1016/j.aiopen.2022.09.001.

[5] I. Kadek Sastrawan, I. P.A. Bayupati, "Detection of fake news using deep learning CNN–RNN based methods," *ICT Express*, vol. 8, no. 3, pp. 396–408, 2022, doi: https://doi.org/10.1016/j.icte.2021.10.003.

[6] P. Ghadekar, M. Tilokchandani, A. Jevrani, S. Dumpala, S. Dass, and N. Shinde, "Prediction and Classification of Biased and Fake News Using NLP and Machine Learning Models," *Adv. Intell. Syst. Comput.*, vol. 1311 AISC, pp. 13–20, 2021, doi: 10.1007/978-981-33-4859-2_2.

[7] Chaowei Zhang, Ashish Gupta, "A computational approach for real-time detection of fake news," *Expert Syst. Appl.*, vol. 221, p. 119656, 2023, [Online]. Available: https://doi.org/10.1016/j.eswa.2023.119656

[8] Halyna Padalko, Vasyl Chomko, "A novel approach to fake news classification using LSTM-based deep learning models," *Front. Big Data*, vol. 6, 2023, doi: https://doi.org/10.3389/fdata.2023.1320800.

[9] "(PDF) A Comprehensive Review on Fake News Detection With Deep Learning." Accessed: Apr. 21, 2026. [Online]. Available: https://www.researchgate.net/publication/356366106_A_Comprehensive_Review_on_Fake_News_Detection_With_Deep_Learning

[10] A. Durić, S. Koprivica, D. Nikolić, and D. Stefanović, "Assessing Reproducibility and Accessibility in Hate Speech and Fake News Detection Datasets: A Literature Review (2023-2024)," *2025 24th Int. Symp. INFOTEH-JAHORINA, INFOTEH 2025 - Proc.*, 2025, doi: 10.1109/INFOTEH64129.2025.10959261.

[11] Emmy Danny Ajik, Georgina N Obunadike, "Fake News Detection Using Optimized CNN and LSTM Techniques," *J. Inf. Syst. Informatics*, vol. 5, no. 3, pp. 1044–1057, 2023, doi: 10.51519/journalisi.v5i3.548.

[12] C. K. Hiramath and G. C. Deshpande, "Fake News Detection Using Deep Learning Techniques," *1st IEEE Int. Conf. Adv. Inf. Technol. ICAIT 2019 - Proc.*, pp. 411–415, Jul. 2019, doi: 10.1109/ICAIT47043.2019.8987258.

[13] Sangita M. Jaybhaye, Vivek Badade, "Fake News Detection using LSTM based deep learning approach," *ITM Web Conf.*, 2023, doi: 10.1051/itmconf/20235603005.

[14] I. Y. Khudhair, S. H. Majeed, A. M. S. Ahmed, M. A. K. Alsaeedi, and F. M. Aswad, "An Improved Hybrid GRU and CNN Models for News Text Classification," *Int. J. Informatics Vis.*, vol. 9, no. 1, pp. 303–313, Jan. 2025, doi: 10.62527/JOIV.9.1.2658.

[15] Al-obaidi, Saja A. and Çağlıkantar, Tuba, "Automated fake news detection system," *Iraqi J. Comput. Sci. Math.*, vol. 5, no. 4, 2024, [Online]. Available: https://ijcsm.researchcommons.org/cgi/viewcontent.cgi?article=1200&context=ijcsm

[16] Mohammed E.Almandouh, Mohammed F. Alrahmawy, Mohamed Eisa, Mohamed Elhoseny & A. S. Tolba, "Ensemble based high performance deep learning models for fake news detection," *Sci. Rep.*, 2024, [Online]. Available: https://www.nature.com/articles/s41598-024-76286-0

[17] J. A. Waqas Haider Bangyal, Rukhma Qasim, Najeeb ur Rehman, Zeeshan Ahmad, Hafsa Dar, Laiqa Rukhsar, Zahra Aman, "Detection of Fake News Text Classification on COVID-19 Using Deep Learning Approaches," *Comput. Math. Methods Med.*, 2021, doi: https://doi.org/10.1155/2021/5514220.

[18] Faisal A. Alshuwaier, Fawaz A. Alsulaiman, "Fake News Detection Using Machine Learning and Deep Learning Algorithms: A Comprehensive Review and Future Perspectives," *Computers*, vol. 14, no. 9, p. 394, 2025, doi: https://doi.org/10.3390/computers14090394.

[19] Nida Aslam, Irfan Ullah Khan, Farah Salem Alotaibi, Lama Abdulaziz Aldaej, Asma Khaled Aldubaikil, "Fake Detect: A Deep Learning Ensemble Model for Fake News Detection," *Complexity*, 2021, [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1155/2021/5557784

[20] Nabeela Kausar, Asghar AliKhan & Mohsin Sattar, "Towards better representation learning using hybrid deep learning model for fake news detection," *Soc. Netw. Anal. Min.*, vol. 12, no. 165, 2022, [Online]. Available: https://link.springer.com/article/10.1007/s13278-022-00986-6

[21] S. Raza and C. Ding, "Fake news detection based on news content and social contexts: a transformer-based approach," *Int. J. Data Sci. Anal.*, vol. 13, no. 4, pp. 335–362, May 2022, doi: 10.1007/S41060-021-00302-Z.

[22] Jia Li, Minglong Lei, "A Brief Survey for Fake News Detection via Deep Learning Models," *Procedia Comput. Sci.*, vol. 214, pp. 1339–1344, 2022, doi: https://doi.org/10.1016/j.procs.2022.11.314.

[23] "fake-and-real-news-dataset." Accessed: Apr. 21, 2026. [Online]. Available: https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset

[24] Nicolas Zucchet, Antonio Orvieto, "Recurrent neural networks: vanishing and exploding gradients are not the end of the story," *arXiv:2405.21064*, 2024, [Online]. Available: https://arxiv.org/abs/2405.21064

[25] V. Kovalenko, I. Dorohyi, and K. Doroshenko, "Comparative analysis of models for fake news detection and classification using gru," *Nauk. Pr. Donec'kogo nacìonal'nogo Teh. unìversitetu. Serìâ "Občislûval'na Teh. ta Avtom. Pr. Donec'kogo nacìonal'nogo Teh. unìversitetu. Serìâ Občislûval'na Teh. ta Avtom.*, pp. 39–57, Nov. 2024, doi: 10.31474/2786-9024/V2I2(34).313834.

[26] I. D. Mienye, T. G. Swart, and G. Obaido, "Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications," *Inf. 2024, Vol. 15, Page 517*, vol. 15, no. 9, p. 517, Aug. 2024, doi: 10.3390/INFO15090517.

[27] Željko Vujović, "Classification Model Evaluation Metrics," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, pp. 599–606, 2021, doi: 10.14569/IJACSA.2021.0120670.

[28] Alaa Tharwat, "Classification assessment methods Open Access," *Appl. Comput. Informatics*, vol. 17, no. 1, pp. 168–192, 2021, doi: https://doi.org/10.1016/j.aci.2018.08.003.

[29] D. Yi, I. Kim, and S. Bu, "VARIATIONS OF TRAINING PROCESS IN VANILLA RECURRENT NEURAL NETWORK FRAMEWORK," *Neural Netw. World*, vol. 34, no. 2, pp. 73–87, 2024, doi: 10.14311/NNW.2024.34.005.

[30] V. U. Gongane, M. V. Munot, and A. D. Anuse, "A survey of explainable AI techniques for detection of fake news and hate speech on social media platforms," *J. Comput. Soc. Sci. 2024 71*, vol. 7, no. 1, pp. 587–623, Mar. 2024, doi: 10.1007/S42001-024-00248-9.