Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Transposed Residual Convolutional BiGRU Network Model for Fake News Detection based on Multi-Attention

Satinder Pal[1*], Amit Jain[2], Anil Kumar Lamba[3]
[1]Research Scholar, Geeta University, Panipat,Haryana-132103, India.
[2]Professor, Chandigarh University, Gharuan, Punjab-140413, India.
[3]Professor, Geeta University, Panipat, Haryana-132103, India.
*Corresponding Author: Satinder Pal
Mail ID: spcm311@gmail.com

ARTICLE INFO

Received: 22 Sept 2024

Revised: 04 Oct 2024

Accepted: 16 Nov 2024

ABSTRACT

The goal of this paper is to spot fake news on X platform (Formerly Twitter) using a special deep learning model called MA-TRC-BGRN. It analyzes data from two sources: the Twitter EDA dataset and the Fake NewsNet dataset, to tell apart real and fake news. Before training, the text is cleaned up by removing extra words, symbols, and irrelevant tags, and by applying steps like stemming, lemmatization, and case folding. The cleaned text is then converted into meaningful features using the NoR_BeRT model. Since the extracted features can be too many and complex, the Dualistic Battle Royale Optimization (Du_BRO) algorithm is applied to pick only the most useful features for better accuracy.

The novel MA-TRC-BGRN model is then proposedto detect Twitter fake news. The Python tool is used to carry out the simulation in this case. Also, the efficiency of the proposed model is improved by fine-tuning the parameters using the Fire Hawks Optimization (FHO) approach. The Mean Absolute Error (MAE) value of the proposed method on the Twitter EDA dataset is 0.003%, and the MAE value on Fake NewsNet is 0.002%, indicating that the proposed method outperforms other existing methods. By employing a hybrid deep learning methodology, the proposed model is more efficient.

Keywords: Fake news detection, Multi-Attention module, Bidirectional Gated Recurrent Unit, Dualistic Battle Royale optimization, Normalization based Bidirectional Encoder Representation from Transformers, Twitter platform.

1. INTRODUCTION

As more people use social media, fake information is spreading around the world. This is a big problem that is getting worse every day as people lose faith in the original accuracy of information. As a result, social media is under pressure to address this issue because the dissemination of misleading information effects public perception of a certain event (Kumar et al. 2020; Mridha et al. 2021). Now a days, information is obtained with a single click on the Internet. There is freedom to share every single story online, which results in more false information being spread day by day. The use of clickbait titles and headlines is increasing with the aim of forwarding inaccurate and unprofessional messages in order to generate advertising revenue (Sharma et al. 2022; Rastogi et al. 2021). Users wonder whether they could alter the original message to spread hot topics or rumours across social media for political or financial gain. Each person's role in news sharing has grown dramatically over time. It is difficult to evaluate the authenticity of information from its source (Konagala & Bano, 2020; Swetha & Priyanka, 2023). As a result, fake news has recently received a lot of attention on a variety of social media platforms, including Facebook, Twitter, and Google.Students struggle to identify originality in internet data, and subpar reporting leads to inaccurate linkages, contexts, and content (Djenouri et al., 2023; Luvembe et al., 2023).Fake news, which causes problems in both the political and technical areas, is severely affected. Facebook has already done what it can to combat misinformation, and it has taken action against the spread of false information (Capuano et al., 2023; Kong et al., 2020).

The main issue here is that the network is active, since it is based on real-time data, and early control of the spread of rumours is required, which is possible by identifying rumours by their origin. One way to stop the spread of rumours is to identify them based on their source of origin (Hiramath & Deshpande, 2019; Kaliyar et al., 2020). Applying an appropriate distribution model helps in investigating the source of fake news. Then, based on the source, either the multi-source approximation-based approach or the single-source rumour-based approach was applied to prevent the spread of rumours in the network (Wani et al. 2021; Nasir et al. 2021). Finding out real and fake news is a challenging approach. Fake news identification is the subject of extensive research, primarily in two areas: deep learning (DL) and natural language processing (NLP). False reports from unapproved sources have been located using a number of techniques (Monti et al. 2019; Aslam et al. 2021). Fake news detection using natural language processing techniques such as Bigram Term Frequency-Inverse Document Frequency (TF-IDF) and probabilistic context-free grammar detection.

Testing is then conducted using a variety of classification techniques, including gradient boosting, stochastic gradient descent, bounded decision trees, random forests, and support vector machines (SVMs) (Choudhary & Arora, 2021; Umer et al., 2020). Subsequently, the CSI model integrates three attributes of false information: the content of an item, the origin, and user response. Deep learning, which is a new technology that extracts high-dimensional features automatically with high accuracy, is an emerging area of research. The DL method includes various approaches, such as the convolutional neural network (Verma et al. 2023), Recurrent Neural Network (Xie et al. 2023), Graph Neural Network, Generative Adversarial Network, Attention Mechanism, Bidirectional Encoder and Ensemble Approach (Luqman et al. 2024). The convolutional neural network consists of convolution, pooling, and regularization layers. The recurrent neural network has a long short-term memory and a gated recurrent unit(Zhang et al. 2020; Bahad et al. 2019). In additional, recent studies use advanced transformer models for detecting fake news from mixed languages and attaining better performances (Guo et al., 2023).

Motivation

Recently, social media has become more popular and has gained wide popularity among numerous people around the world due to its low cost, rapid distribution and effortless access. Nowadays, the sharing of fake news on Twitter is increasing and is one of the hottest research topics in information technologies. Such fake information can create social harm by altering public discourse and influencing political processes such as policy making and elections. Thus, an accurate detection of fake news is important to mitigate such consequences and promote the reliability of shared information. Analyzing fake news on Twitter can protect the integrity of several democratic institutions and promote a safer society. Also, the automatic detection of fake news can secure the users from exploitation, manipulation and harm. Moreover, the motivation behind the fake news detection is to enhance the integrity of shared information on Twitter and promote public safety. The existing detection models have experienced various limitations in fake news detection tasks. Thus, to address the drawbacks of the existing techniques and thereby attain superior performance while detecting twitter fake news, the proposed study introduced a novel deep learning model in this work.

The major objectives of the proposed work are,

- To remove the irrelevant information behind the text, pre-processing is done in the initial step using popular natural language processing (NLP) methods.
- To achieve higher classification accuracy, the most important features behind the text are identified using a new NoR_BERT model.
- To overcome the problem of feature dimensionality, the optimal features are chosen by introducing the Du_BRU algorithm, thus reducing the computational complexity problem.
- To propose a novel MA-TRC-BGRN model for detecting available fake information behind the input tweet data with reduced complexity.
- To validate the suggested model's performance, various metrics are examined, and the achieved performance is compared to other current approaches for proving the efficacy of the proposed work.

The rest of the paper is organized as follows: Section 2 covers the relevant literature, Section 3 describes the proposed methodology for detecting fake news on the Twitter platform, Section 4 deals with the results and discussion of the proposed method with several existing methodologies, and Section 5 presents the conclusion of this work with an

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

effective future scope.

2. RELATED WORKS

The news industry has shifted focus from traditional print media to online platforms like Twitter and Facebook in recent years. The proliferation of false information via social media is a global problem. Deep learning techniques can be used to address this issue. Combining a deep convolutional neural network with a bidirectional encoder, prototypical transformers (BERT) were first suggested by Kaliyar et al. (Kaliyar et al. 2021). BERT is a pre-trained model built on transformer encoders. It learns sentence meaning by masking random words and predicting them, which helps capture context in both directions. It can learn from plain, unlabeled text, but it needs huge training data, high computing power, and is costly because of its complexity. Using a deep learning technique called Long Short-Term Memory (LSTM) neural networks, Chauhan et al. (Chauhan & Palivela, 2021) developed a system to identify false news articles such as Crime, rumours, accidents, and humorous news are the target audience for this piece. In addition, word embedding was used for vector representation of documentary words using neural networks. The tokenization technique and the N-gram concept used here are used for feature extraction. In place of a neural network, LSTM employs cell blocks. Such cells include input, output, and a forget gate. The model's precision was increased by including stop words in pre-training data processing. Although it has various benefits, there are some drawbacks, such as complexity and unsuitability for online learning. It is critical in e-commerce apps to detect fake news online.

Although several approaches present challenges in terms of processing speed, Zhang et al. (Zhang et al. 2023) proposed a new deep learning model based on rapid fake news detection for cyberphysical social services in cyberspace. A novel model based on a convolutional neural network was developed to solve this problem with the Chinese language being treated as objective, news being short text with keywords, and features being extracted by convolution to ensure processing speed and recognition capacity. Results are then evaluated using the data sets, which are the rumour and CHEF datasets that rapidly detect fake news through automatic classification. This method has the disadvantage of making it difficult to detect the location and orientation of fake news and requires a large number of training data sets.

As social media usage increases rapidly, it leads to misinformation and disinformation for political, unfair, and financial reasons. Rastogi et al. (Rastogi et al. 2021) introduced a hybrid HyproBert method for automatically detecting fake news. Tokenization and word embedding are handled by DistilBERT in HyproBert. The spatial features are extracted by embedding, then sent to BiGRU to be paired with contextual information, and finally, the combined output of BiGRU and CapsNet is passed on to the self-attention layer to determine the spatial relationship between the features. Using the two data sets ISOT and FAKES, the results are evaluated. Because of the complex structure, training can be time-consuming, expensive, and difficult.

The detection of fake information on social media was addressed by Yildirim and Gungor et al. (Yildirim, 2023), who presented a hybrid metaheuristic multi-threaded technique. It uses super-threads for concurrent monitoring and parallel search patterns, allows easy exploration of attribute combinations, and also allows software models to be easily implemented. For fraud detection, it employs a parallel hybrid optimization strategy. Three data sets containing information on COVID-19 and daily politics are used to conduct the review. However, the resulting system's performance is impacted by the local maxima problem inherent in the optimization approach.

Ni et al. (Ni et al. 2021) presented a framework using a neural network model to spot bogus news in the 140-character tweets that were provided. In order to better understand the world from multiple perspectives, a novel multi-view attention network (MVAN) was created. As a result of attention models, the neural network was able to adequately take in signals and data from both the input source and the propagation structure. In addition, the mechanisms of attention are ensured that developed network was able to analyze keywords from entered messages and fraudulent users from the distribution system. Comparing the state-of-the-art with two distinct real-world datasets, such Twitter 15 and Twitter 16,, the model achieved superior results.

Hakak et al. (Hakak et al. 2021) used a new methodology based on machine learning. These previous investigations involved pre-processing, feature selection, and optimal classification. A decision tree, random forest, and extra-tree algorithms were used in this study to make precise classifications. The random search hyperparameter optimization approach was used for parameter optimization to obtain good performance. On the Liar and ISOT datasets, the created model outperformed other techniques in terms of accuracy.

The DL-based technique for detecting fake information in social media was defined by Ali et al. (Ali et al. 2022). The main components of this study are three. The first step was to extract features from message content and preprocess them using the Natural Language Processing (NLP) method, followed by word augmentation. employing N-gram, followed by the sequential DL (SDL) approach to extract hidden features, and lastly a multilayer perceptron to classify them. LIAR and ISOT gathered the dataset at around 12.8 K and 44.8 K, respectively.The experiment's accuracy was 100% and its precision was 99.94%. Nonetheless, this approach is exclusively used for news content in English and cannot be used to handle image-based material.

For Twitter fake news identification, Hamdi et al. (Hamdi et al., 2020) proposed a hybrid technique that makes use of user characteristics and graph embeddings. This current work employs a methodology that allows one to evaluate the statistical sources in relation to the reliability of Twitter. This approach makes use of node2vec to get the characteristics out of the Twitter follower's graph. This method starts by looking for an acceptable Twitter social graph, as seen in the embedded map. The search agent then collects the topics they tweeted about for each customer in the graph. Next, link each user's credibility to a static threshold to indicate whether or not a customer is an SOFN. A graph embedding operation removes the node function when the relevant graph is specified. This method fails to give accurate results for fake news prediction on Twitter.

Zervopoulos et al. (Zervopoulos et al. 2020) developed a Hong Kong protest to detect fake news on Twitter using NLP. In this present work, the Chinese tweets, which are reserved alongside originally English tweets, are decoded into English from the relevant data set. The relevant tweets are determined through a language-independent screening process. Renowned ML algorithms are used to categorize the tweets characterized by a feature value vector from the removed, nominated and pre-processed datasets, largely revolving around linguistic usage, with word entropy being a new feature. However, this method is unsuitable for detecting other languages and only predicts the Chinese language in English tweets.

Problem statement

With the rapid expansion of social media platforms, the propagation of disinformation and fake news has become increasingly prevalent. Fake news can be disseminated in several forms: images, text, audio and video recordings. Nowadays, Twitter is becoming more popular because it aids users in sharing short messages termed as tweets with their followers, and this is considered a powerful platform for communication around the world. However, due to the large number of users, the sharing of fake news on Twitter is increasing, which has a high impact on society. In order to automatically detect fake news on the Twitter platform, it is necessary to develop an advanced technique. In recent years, several studies have developed varied approaches for detecting fake news on Twitter. Nevertheless, the existing techniques have faced several issues while detecting fake information because of their inefficiency. Learning hierarchical representation is critical for correctly recognizing fake information embedded in input tweets.However, the existing techniques failed to fetch the needed information, and hence the accuracy is limited. The overall performance suffers as a result of the ineffective usage of features in the absence of target class redundancy and increased internal similarity.Also, some of the existing techniques are unable to handle large amounts of information and lead to overfitting problems. In addition, the complex training process of the existing detection models generates higher computational complexity issues. Thus, to overcome the mentioned issues in the previous methods, the proposed study utilizes an MA-TRC-BGRN model for accurately detecting fake news on the Twitter platform.

3. PROPOSED METHODOLOGY

Detecting fake news on social media is a major research focus, as it helps reduce the harmful impact of misinformation. However, this task is challenging since it requires analyzing large amounts of information from online content. To address this, the present study introduces a new method for detecting fake news using the selected dataset. Figure 1 presents the block diagram of the proposed approach.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

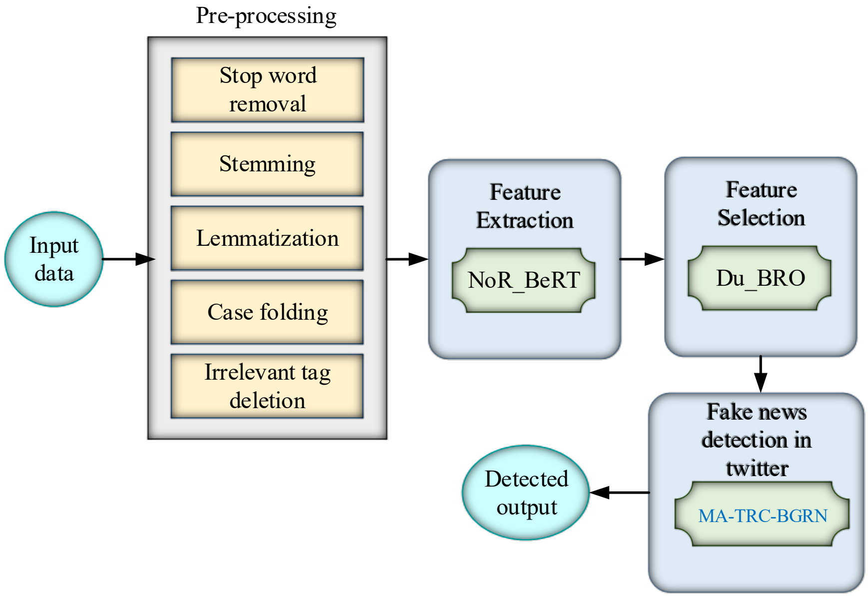

Figure 1: Block diagram of the proposed work

Data gathering, feature extraction, feature selection, and fake news detection are the several phases of fake news detection. To enable reliable false news identification, a unique feature selection strategy is developed to address dimensionality difficulties, enabling greater classification accuracy with less time consumption.

3.1 Pre-processing

Data pre-processing is the process of changing raw data into a flawless data set. The text data needed to run the false news detection algorithm is initially taken from the Twitter EDA dataset. To eliminate undesired noise in the data and improve detection performance, the obtained data are pre-processed utilizing stop word removal, stemming, lemmatization, case folding, and irrelevant tag elimination (Nawaz et al., 2023).

3.1.1 Stop word removal

There are several advantages to eliminating the stop word. Firstly, it reduces storage overhead due to the reduction of large amounts of text data. Second, it reduces noise by removing stop words so the focus is turned on other important substances. Stop word removal also removes stray letters and coded characters.

3.1.2 Stemming

Stemming on of the technique that determines the unwanted concept of a word. It is employed to transform a word's form, thereby creating the fundamental concept that aligns with the structure of a refined morphology. Stemming involves removing suffixes such as "-s," "-ed," "-mis," "-ize," and "-de" from words. For example, words like "programs," "programming," and "programmer" can all be reduced to the common root "program." This technique helps to condense data entries within a network, decreasing both complexity and size.

3.1.3 Lemmatization

Lemmatization in linguistics refers to the process of consolidating the various forms of a word into a singular entity known as its base form or vocabulary form. This technique serves as a method of standardizing text within the realm of natural language processing. To remove the affixes from words, lemmatization employs both morphological and terminological analysis. For instance, the phrase "Building has floors" is simplified to "Buildings have floors." The stemming method is also faster than lemmatization, but the texts may end up using non-vocabulary and irrational words.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

3.1.4 Case folding

Case folding refers to a technique utilized on a sequence of characters that transforms recognized lowercase letters into uppercase letters. In the indexing procedure, the case folding method is employed to facilitate comparisons. For varying each character, transitioning to lowercase case folding is the most effective stage. Subsequently, characters are removed by discarding all unnecessary characters, specifically all elements other than lowercase letters (a-z). A notable instance of case folding is that the German lowercase letter 'ß' corresponds to 'ss'.

3.1.5 Irrelevant tag deletion

The process of eliminating irrelevant tags focuses on removing unnecessary tags from the text document. Common examples of such tags include HTML tags, PHP tags, JavaScript code, one-time commands, and style sheets. The procedure for deleting irrelevant tags will involve rearranging a series of new lines, followed by a shift towards acceptable tags.

3.2 Feature Extraction

Feature extraction transforms text into numerical features that can be analyzed while maintaining the statistical properties of the original dataset. To enhance classification accuracy, contextual text features are derived from the preprocessed data through the Normalization-Based Bidirectional Encoder Representations from Transformers (NoR_BeRT) model.

3.2.1 BERT model

The BERT architecture is composed of an embedding layer followed by multiple bidirectional transformer encoders. The embedding layer converts input text sequences into vector representations. Each transformer encoder then refines these vectors using a combination of multi-head self-attention and a feed-forward neural network, passing the outputs sequentially through the stack. The self-attention mechanism captures bidirectional contextual information from the text, while the feed-forward network extracts hierarchical features. To address issues such as exploding or vanishing gradients, both the self-attention and feed-forward components are equipped with residual connections and normalization layers.

3.2.2 Feature Normalization

Count-based features are taken as n-grams, but their values need normalization to avoid being influenced by text length. Classification accuracy drops when global weighting methods like TF-IDF or loss entropy are applied. These weighting schemes are less effective for fake news detection because repeated n-grams often carry useful signals, but their importance gets reduced by global normalization. The proposed study used the L2 norm for normalizing the features, where the L2 norm weighting system is used for the character bigram, which helps attain the usefulness of varied character sequences in the input text. The L2 norm normalization is generally utilized in text-processing tasks, including information retrieval, document clustering and text classification.

The dual weighting system of a feature constrains values and has only dualconceivable values; if the n-gram is present in the text, it will be , otherwise, it will be . It is represented as,

$$\begin{cases} g_b = 1, & if\ lrb > 0 \\ g_b = 0, & if\ lrb = 0 \end{cases} \quad (1)$$

Where lrb is defined as the number of times that the term performs in text.

After evaluating the weights of all character bigrams, the L2 norm is applied to normalize the weights. The L2 norm is dependent on the Euclidean distance for normalization:

$$\|m\| = \sqrt{m^B m} = \sqrt{\sum_{b=1}^{t} m_b^2} \quad (2)$$

In the L2 norm weighting process, the character bigram's weight is evaluated as the square root of its vector count. This computation helps to maintain the corresponding importance of all character bigrams. For any vector representation,

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

the L2 norm unit represents m $\|m\| = 1$, denotes the vector's length. Therefore, the vector m can be normalized as $\frac{m}{\|m\|}$. The L2 norm frequency weighting system measures the feature vector representation of all data points to obtain the length of a unit norm.

3.3 Feature selection

It will take time and create a dimensional problem if all of the collected features are directly provided in the classification procedure. To reduce this problem, a feature selection phase is included in this work. The Dualistic Battle Royale Optimization (Du_BRO) algorithm is used to pick the best features from the extracted set, helping to solve the problem of high feature dimensions.

3.3.1 Battle Royale Optimizer algorithm (BRO)

Battle Royale games inspired the Battle Royale Optimizer algorithm. The major candidate solutions in the issue space are randomly distributed in the battle royale games. Subsequently, when each solution is related to its adjacent neighbour, the improved solution can be called an active search agent and the other a lazy search agent in terms of fitness value. The loss degree of each candidate solution is preserved as a parameter that can be increased after each loss in each candidate solution. If a solution is compromised repeatedly for a threshold duration of 3 to 6, depending on the problem, it will be solved and rearranged using the following equation (3),

$$m_{loss}^{b+1} = q\left(wh_t - dh_t\right) + dh_t \tag{3}$$

According to the following equation (4), reordering will be effective if its loss level does not exceed the threshold,

$$m_{loss,t}^{b+1} = m_{loss,t}^{b} + q\left(m_{finest,t} - m_{loss}^{b}, t\right) \tag{4}$$

Where q is denoted as an arbitrarily produced amount based on uniform dissemination in the range $[0,1]$. The position of the impaired and finest solution is signified as $m_{loss,t}$ and $m_{finest,t}$ in the dimension t. The inferior and superior confines of dimension in problem space are represented as dh_t, wh_t

Each rearrangement starts over with the rival solution in order to get closer to the optimal solution, which is still a long way off. The main thing about this algorithm's search process is that the safe area in the problem gets smaller with each iteration as it moves toward the best solution by equation (5),

$$\Gamma = \Gamma + round\left(\frac{\Gamma}{2}\right), if\ b \geq \Gamma \tag{5}$$

When the reaches the rate of Γ, the region shrinks down Γ. has been modified by,

$$\Gamma = \frac{Maxcenter}{round\left(\log_{10}\left(Maxcenter\right)\right)} \tag{6}$$

Where $Maxcenter$ denotes the maximum amount of generations. The main perception of the space constraints is to transfer the entire solution of the participants towards the best possible solution. To support elitism, the best solution is saved in each iteration. The problem space can be reduced using equation (7),

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

$$\begin{cases} dh_t = M_{finest,t} - RH(m_t) \\ wh_t = M_{finest,t} + RH(m_t) \end{cases} \quad (7)$$

Where,the standard deviation of the entireinhabitants in dimension t is signified as $RH(m_t)$. Table 1 represents the contingency table for bit evaluation.

Table 1: Contingency table among two dualistic variables for bit evaluation.

BIT	0	1
0	$\#of\,00$	$\#of\,01$
1	$\#of\,10$	$\#of\,11$

3.3.2 Proposed approach: Dualistic Battle Royale Optimization algorithm (Du_BRO)

In the Du_BRO algorithm, search agents start with random positions, spread evenly across the search space. Based on the fitness function which was specified in equation (8), the optimal features are nominated.

$$f_\vartheta = \varphi * \psi + (1-\varphi) * \frac{|K_c|}{K_d} \quad (8)$$

Where, the error value is denoted as ψ, φ signifies the parameter that affects the data, the total amount of data obtained in the dataset is indicated as K_d and the total amount of data from the extracted features is represented as K_c.

If the produced arbitrary amount is lesser than , the rate is allocated to a specifiedability , or else . The number of solutions is equivalent to the number of abilities. Every solution is a t-dimensional dualistic vector,

$$M_b = (m_{b1}, m_{b2}, \cdots, m_{bt}) \quad (9)$$

In different problems, this vector can be interpreted differently; for example, in the Uncapacitated Facility Location Problem (UFLP), m_{bs} match up to the s^{th} ability of the b^{th} solution. Then, for the predetermined amount of iterations, this algorithm is reiterated, here *Max iter* is a set of Max-cut and UFLP problems. In every iteration, all search agent is compared with their adjacent neighbour. Fordualistic variables, the adjacent neighbour is determined based on the comparison measures related to the following equation (24);

$$com(M_b, M_s) = \frac{\#of\,00}{\#of\,00 + \#of\,01 + \#of\,10} \quad (10)$$

From the above equation, the number of bits is $\#of\,00, \#of\,01, \#of\,10\, and\, \#of\,11$ correspondingly, which includes the values as $"00", "01", "10", and\, "11"$ in the dual neighbour solution. The variable $\#of\,11$, in equation (24), the amount of bits with the value in each search agent, is unnoticed. The perception behind this problem is that in unbalanceddualistic variables, the amount is commonlybetter than $\#of\,00, \#of\,01, \#of\,10$, and totalling it to the denominator of the equation (24)couldunreasonablylessen the comparison score.After that, when a search agent is related to its adjacent neighbour, the solution with superior fitness value, as determined by the fitness function, is good, and the other is phony. Du_BRO controls the active and lazy search agents using different approaches. A mutation job is permanently applied to the active search agent and must modify the currently active search agent's position in order to make it the last better active search agent. In the mutation task, only two arbitrarily selected bits of the active search

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

agent solution would be flipped.

If the Du_BRO has not achieved the maximum amount of losses, it executes a crossover job on the lazy search agent solution. The crossover task has been employed in three different ways: single-point crossover, dual-point crossover, and uniform crossover. In the single point, dual point, and uniform crossover ways, the search agent solution is arbitrarily divided into dual, three, and t portions. If the lazy search agent suffers the maximum number of losses, it will be re-spawned later as an important mutation task. The drowsy search agent is re-spawned in another permissible place of the searching space, just as it was throughout the searching process. This necessary mutation is dependent on the current iteration. Du_BRO reverses the β bits of the slow search agent solution from $1\, to\, 0$, resulting in the equation (11) below.

$$\beta = dn - (dn-1) * \frac{C-iter}{M-iter} \qquad (11)$$

Where the maximum iteration amount is denoted as $M-iter$, the current iteration amount is represented as $C-iter$, and dn signifies the number of dimensions bits in every solution. The core idea is that similar to the Battle Royale game, Du_BRO constrains the search agent's solution to find the best solution by decreasing the search space, where the solutions compete with each other. This task is repeated until either the maximum number of iterations is achieved or a valid solution is found. The 2-bit mutation is tested for the optimal solution after each solution modification for the Lazy Search Agent or Active Search Agent. If the optimal solution achieves a higher fitness value after the mutation job, it is modified; otherwise, it remains unchanged.Thus, the total amount of mutation tasks on the finest solution in the course of the entire algorithm would be $M-iter*a$, where the amount of the entire solution is denoted as a.

3.4 Detection

The fake news can be detected effectively through a DL-based approach called the MA-TRC-BGRN model. Figure 2 denotes the block diagram ofthe proposed MA-TRC-BGRN for detecting fake news on the Twitter platform.

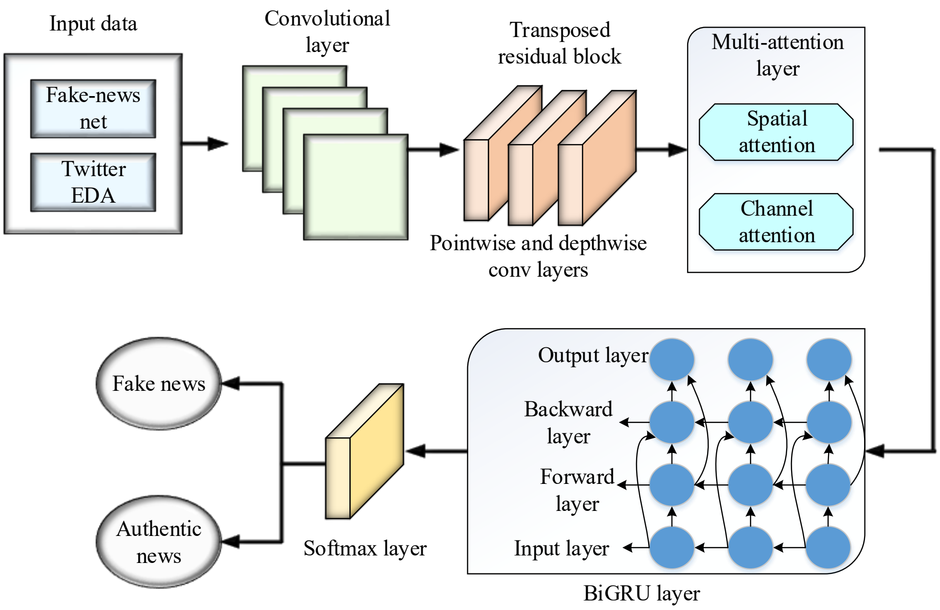

Figure 2. Block diagram of MA-TRC-BGRN

The proposed model is the combination of improved CNN and BiGRU networks. In comparison to the standard Bi-directional Gated Recurrent Unit (BiGRU), Convolutional Bi-directional Gated Recurrent Unit (Conv-BiGRU), and Attention Convolutional Network (Attn-Conv), the proposed MA-TRC-BGRN is extremely effective.Because, along with the proposed model, multi-attention is used to attain rich feature information in this work.The input data is first

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

transmitted to the convolution layer, in which the meaningful features are learned, and the data from the convolution layers are fed to thetransposed residual blocks. By using this block, the computational cost is reduced because the parameters are limited. The outcome of residual blocks is then passed to the multi-attention network, which helps reduce the large dimensionality issues. The learned informationis then enteredinto the Bi-GRU layer. Finally, the softmax layer produces the output as fake or authentic news behind the input text data.However, the preceding schemes cannot include the transposed residual blocks for enhancing the convolution operation. Also, the existing method cannot incorporate the dual attention mechanism with BiGRU. Thus, the novelty of the proposed MA-TRC-BGRN model is enhanced as compared with other conventional methods.

3.4.1 Convolutional Neural Network (CNN)

In the computer vision field, CNN is widely used to fetch higher-level feature information and is supportable for several tasks like object detection, segmentation, intrusion detection, etc. The CNN approach is more effective at learning hierarchical representations from raw data and provides useful information. The CNN proceeds TX_q and revenues a feature map of which comprises the supreme standards of the features. For the mining of literal features, the classifier feats the convolution layers. By the performance of the convolution on the equivalent layer, every filter increases the feature vectors and makes the feature map for the succeeding layer. The first convolution layer receives a term TX_b from TX_q of a f -dimensional vector. Assume $M_b \in M^f$ be the f -dimensional vector equivalent to TX_b and $M_{b:b+s}$ is the concatenation of feature vectors $m_b - M_{b+s}$. A $g \in M^{tf}$ filter is smeared to a space of t words, making a novel feature map k_b which can be defined as:

$$k_b = f\left(g \cdot TX_{b,b+t-1}\right) \tag{12}$$

Where the bias is denoted as t, and the tangent non-linear function is signified as f. The filter makes a feature map k by the specified space of features, which is shown in the below equation as follows;

$$k = \langle k_1, k_2, \cdots, k_b \rangle \tag{13}$$

Although CNN learns important feature information, the computational complexity issue is becoming a difficult problem. Thus, to avoid this issue, the proposed study introduced a transposed residual block for enhancing the computational efficiency by diminishing the parameters and computation. The major motive of the transposed residual block is performing point-wise and depth-wise convolution, which can mitigate the computation cost by limiting the parameters. The computation cost of the utilized depth-wise convolution operation is the addition of the evaluation number of deep convolution and 1*1 convolution. The mathematical expression of computation cost in depth-wise separable convolution is given as,

$$D_c = L \times L \times D_{in} \times N_{out} \times N_{out} + D_{in} \times N_{out} \times N_{out} \times D_{out} \tag{14}$$

The evaluation of depth-wise separable parameters is specified as,

$$P_c = L \times L \times D_{in} + 1 \times 1 \times D_{in} \times D_{out} \tag{15}$$

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Where, $N_{in} \times N_{in} \times D_{in}$ denotes the input feature size, $N_{out} \times N_{out} \times D_{out}$ represents the output feature size and $L \times L$ represents the convolutional kernel size. The evaluation of the computational cost ratio of new depth-wise convolution operation to conventional CNN is represented as,

$$c_r = \frac{D_c}{D_{conventional}} = \frac{1}{D_{out}} + \frac{1}{L^2} \tag{16}$$

Where, the ratio of computational cost is represented as c_r.

The mathematical expression of the parameter cost ratio of new depth-wise separable convolution to conventional CNN is mentioned as,

$$p_r = \frac{P_c}{P_{conventional}} = \frac{1}{D_{out}} + \frac{1}{L^2} \tag{17}$$

From Equations (16) and (17), the significance of depth separable convolution is exhibited by varying the multiplication process into multiplication and addition processes. Thus, it can diminish the parameters and cost of computation. Because of the computational characteristics of depth-wise convolution, it cannot swap the number of channels individually. Thus, 1*1 point-wise convolution is included before applyingdepth-wise convolution in a transposed residual block. Through the point-wise convolution mechanism, the dimension gets increased, which leads to attaining better features from a higher dimensional space and aids in promoting network efficiency. The module's input channel is represented as D_{input}, the amount of output channels after applying point convolution is specified as $n \times D_{input}$and the expansion factor is signified as m. After diminishing the dimension, the significant features can be focussed on the channel. Then, the linear activation function is used after reducing the dimension. Figure 3 shows the structure of the transposed residual module.

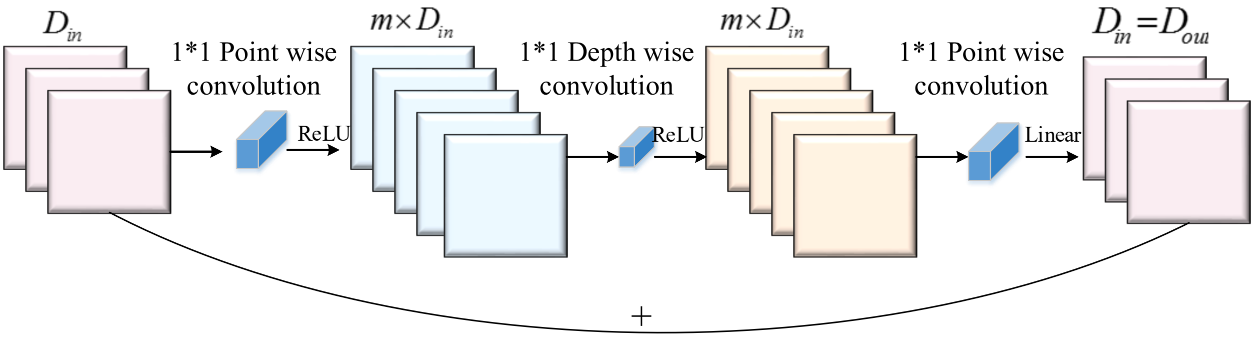

Figure 3 Transposed residual module

3.4.2 Multi-attention layer (MA)

The multi-attention module needs to make the l^{th} area in the explosion specified an in-between feature map $X^d \in N^{L \times G \times T}$ as input $X^d = CNN\left(X^{d-1}\right)$ and also the final concealed state of sentence LSTM $e_{sent}^{l-1} \in N^H$, here the dimension of the concealed state is H. Estimate the weight of the channel attention by the below equation (18);

$$\alpha^d = I_k\left(e_{sent}^{l-1}, X^d\right) \tag{18}$$

Where, the channel attention function is represented as I_k. Next, estimate spatial attention weight β^d by the below equation (19);

$$\beta^d = I_p\left(e_{sent}^{l-1}, I_k\left(X^d, \alpha^d\right)\right) \tag{19}$$

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Where, the spatial attention function is denoted as I_p. Next, with the channel attention weights α^d and spatial attention weights β^d, estimate the refined feature map M^d based on α^d, β^d and X^d;

$$M^d = y\left(M^d, \beta^d, \alpha^d\right) \qquad (20)$$

Where, the linear weighting function is denoted as $y(\cdot)$, that applies the element-wise multiplication. Figure 4 shows the structure of the multi-attention layer.

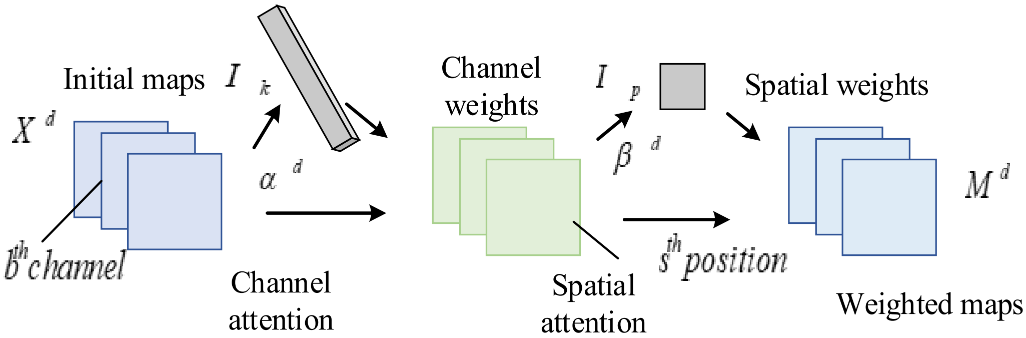

Figure 4: Block diagram of Multi-Attention layer

3.4.3 Bidirectional-Gated Recurrent Unit (Bi-GRU) network

The Gated Recurrent Unit (GRU) neural network is a spherical network architecture that generates current output statistics depending on current input data and previous output statistics. As a result, the output statistics of the GRU neural network at each moment in time are dependent on previous information. As a result, its chain aspect is directly tied to the problem of sequential labelling.

The procedure of producing a linear sum is similar to that of an LSTM cell under the previously calculated and existing parameters. However, the GRU lacks technologies that allow it to view the full state at any time in order to control the level of coverage of its state. The GRU's clever equation is as follows:

$$i_l = \sigma\left(g_i \times \left[e_{l-1}, m_l\right]\right) \qquad (21)$$

$$q_l = \sigma\left(g_q \times \left[e_{l-1}, m_l\right]\right) \qquad (22)$$

$$\tilde{e}_l = \tan e\left(g \times \left[q_1 * e_{l-1}, m_l\right]\right) \qquad (23)$$

$$e_l = \left(1 - i_l\right) * e_{l-1} + i_l * e_l \qquad (24)$$

Where g is the parameter matrix, m denotes the input vector, the output vector is signified as e, i is the update gate state, and the reset gate state is represented as q, correspondingly. In addition, el is a weight matrix, indicating that the current state and past input are used to activate the competitor. Finally, the outcome of the secret unit includes both the updated and the original material. σ stands for the activation function with a value between $0\, to\, 1$. Nodes in a neural network are either ignored or activated by activation functions, depending on the importance of the data they have received.

The next step, following the Bi-GRU neural network feature extraction, is data categorization. This technique uses a softmax activation function to process the result value. The softmax activation function was then implemented at the output layers to determine the final classification decision. Here is the deceptive equation:

$$Q = \arg\max_{v} n\left(u = v \middle| m\right) \qquad (25)$$

$$Q = \arg\max_{v} \frac{\exp\left(output_l\right)}{\sum_{f=1}^{F} output_l} \qquad (26)$$

Where the class label is indicated as k, m, u are the sample feature and the label variable, F is the total amount of classes, and the output of every layer is $output_l$. Since there is a chance, it is clear that softmax classification is a classification result. The end result of Softmax is the chance that some data can be recognized as each class in the current data, as well as the overall chance of the classification result for a subset of the data.

When making the neural network model, it's important to reflect on whether or not the data is overfitting. Regularization is used in this product to get rid of data overfitting. To describe the difficulty of the model, a new pointer is added to the loss function. Here, L1 and L2 are the two regular expressions.

$$L1 = L_0 + \frac{\varphi}{a}\sum_g |g| \qquad (27)$$

$$L2 = L_0 + \frac{\varphi}{2a}\sum_g g^2 \qquad (28)$$

Where the complexity of the model is designated as $\sum|g|\sum g^2$, is the ratio of the model's difficulty loss in the entire loss, and the loss function is indicated as L_0. Nonetheless, in the regularization method, the basic concept is to control the size of the weights and eliminate the faulty acting weight parameters so that the model does not arbitrarily suit the incorrect feature statistics. The weight is condensed () by a constant while utilizing L1 regularity, and L2 is condensed by an amount relational to g . Although the proposed model has a higher learning ability, the loss function in the network can impact the classification performance. Here, the mentioned loss in Equations (27) and (28) is minimized by fine-tuning the parameters of the network model using the FHO approach. The proposed detection model's hyperparameters are as follows: total epoch 300, batch sizes 32, learning rate 0.001, kernel size 3, convolutional filters 128, activation softmax and GRU hidden units 128.FHO is one of the meta-heuristic optimization approachesinspired by the behaviour of black kites, whistling kites and brown falcons. This optimization approach reduces the losses in the network based on the following fitness calculation.

$$Fitness = Min[L1 + L2] \qquad (29)$$

In each iteration, the fitness value is calculated, and the best iteration is selected by attaining a reduced loss value. Initially, the parameters in the network are initialized in the FHO approach for analyzing the initial positions of vectors in the search space. The position updation of search agents for attaining optimal parameter values in FHO is represented as,

$$SA_1^{new} = SA_1 + (rand_1 \times A - rand_2 \times SA_{near}), h = 1,2,......m \qquad (30)$$

Where, SA_1^{new} mentions the current position vector of h^{th} search agent, A represents the global best solution in the specific search space, SA_{near} specifies another search agent and $rand_1$ and $rand_2$ represents the random numbers that are uniformly distributed in the range of 0 to 1. Thus, based on Equation (44), the parameters are fine-tuned in the proposed network model using the FHO approach. Thus, the proposed study effectively detects fake news from the inputs through a novel MA-TRC-BGRN model.

4. RESULTS AND DISCUSSION

This section presents the results and analysis of the proposed technique. Here, the simulation is performed using the Python tool. The input data are taken from the Twitter EDA dataset and the FakeNewsNet dataset. These two datasets involve tweets in text format. The EDA is also called data exploration; EDA is a phase in the data analysis process, and several methods are enhanced to utilize and understand the dataset.This dataset is highly concentrated on Twitter and contains several tweets attained from the Twitter platform. The Twitter EDA dataset is widely used by researchers, analysts and data scientists to explore and detect several aspects of Twitter activities. In this work, the Twitter EDA dataset is gathered from the Kaggle repository, and it involves a total 7613 number of data,of which 4567 data are used for training and 3046 data are used for testing. On the other hand, the FakeNewsNet dataset is attained from Github, which involves a total 22140 number of samples in which 13284 data are used for training and 8856 for testing.This

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

FakeNewsNet dataset is developed to support researchers' interest in analyzing fake information on social media platforms. Because of the large number of input samples, this dataset is widely used to prove the effectiveness of developed detection models. For training and testing purposes, the datasets are split into a 60:40 ratio. Also, the classes involved in these two datasets are real and fake, which aids in performing binary classification. The proposed method was experimentally analyzed by validating the performance parameters that are discussed below.

4.1 Evaluation metrics

The proposed technique considers the following performance metrics: accuracy, precision, recall, F1 score, Kappa, MAE, MSE, and RMSE.

- **Accuracy:** Accuracy is the degree to which the observed object value closely matches the actual object value. The precision equation is shown below.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{31}$$

- **Precision:** Precision is defined as the degree to which frequent dimensions in identical conditions show identical results. The equation to evaluate precision is given below:

$$\Pr ecision = \frac{TP}{(TP+FP)} \tag{32}$$

- **Recall:** Recall is defined as the proportion of correctly identified positive samples relative to the total number of positive models. It is derived using the following equations,

$$\operatorname{Re} call = \frac{TP}{(TP+FN)} \tag{33}$$

- **F1 Score:** The F1 Score is defined as the choral mean of a system's precision and recall values. It can be evaluated by the subsequent equation:

$$F_1\ Score = 2 \times \frac{(\operatorname{Re} call \times \Pr ecision)}{(\operatorname{Re} call + \Pr ecision)} \tag{34}$$

- **Kappa:** Kappa is defined as a relationship to measure the unity of two or more types of outcomes using two or more approaches. It is expressed mathematically in the following equation,

$$\mathrm{K} = \frac{Q_{observed} - Q_{chance}}{1 - P_{chance}} \tag{35}$$

- ***MAE:*** Mean Absolute Error (MAE) is obtained by finding the mean of the sum of total variation among the detected value and exact value.

$$MAE = \frac{1}{m}\sum_{i-1}^{m}\left|b - \hat{b}\right| \tag{36}$$

- ***MSE:*** The mean square error (MSE) is a measure of the error squares, which is the squared difference between the expected and actual values. It is mathematically represented by the following equation.

$$MSE = \frac{1}{m}\sum_{i=1}^{m}\left(b_i - \hat{b}_i\right)^2 \tag{37}$$

- ***RMSE:*** Root Mean Square Error (RMSE) can be defined as a measure of the difference between the scores detected by a model and the exact values. RMSE is mathematically expressed in the following equation.

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left|b - \hat{b}\right|}^{2} \tag{38}$$

Here, the exact value is denoted as band $\hat{b}$ and gives the attained value and m gives the number of samples for

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

evaluation. Denotes the True Negative value, TP specifies the True Positive value, FP represents the False Positive value and FN gives the False Negative value.

4.2 Performance and Comparison Analysis

Using the Twitter EDA dataset and the Fake News Net dataset, this section discusses the results and analysis of the proposed method, as well as a comparison of the proposed technique with existing techniques. Additionally, the proposed model is contrasted to existing techniques such as BiGru (Bidirectional Gated Recurrent Unit), Attn-Conv (Attention Convolutional) and Conv-BiGru (Convolutional Bidirectional Gated Recurrent Unit).

- **Twitter EDA dataset:** Figure 5 compares the suggested method's training and testing accuracy arcs by altering the epoch values.

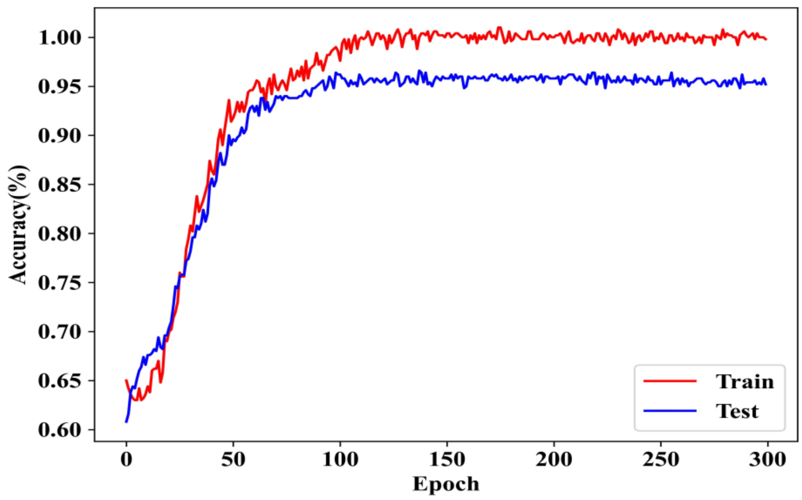

Figure 5 Training and testing accuracy

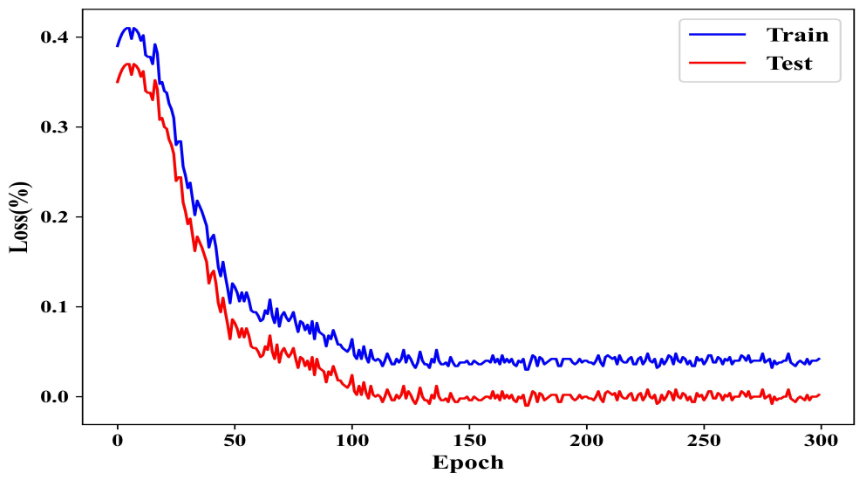

Figure 6: Training and testing loss

This figure shows that accuracy increases when the epoch value is close to zero and reaches the peak of training and testing accuracy of more than 0.90% when the epoch value is just above 50. After that, the accuracy curve remains stable values for the subsequent epoch. At higher epoch values, the proposed method shows higher accuracy for training and testing. Figure 6 shows the training and testing loss arcs of the proposed method by varying the epoch values.

This figure shows that the loss begins to decrease when the epoch value is just above zero and reaches the bottom of training and testing loss of less than 0.05% when the epoch value is just above 50. After that, the loss curve for success

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

remains stable epoch values. For higher epoch values, the proposed method shows minimal losses for training and testing. Figure 7 compares the proposed method's accuracy to existing approaches by altering the learning rate.

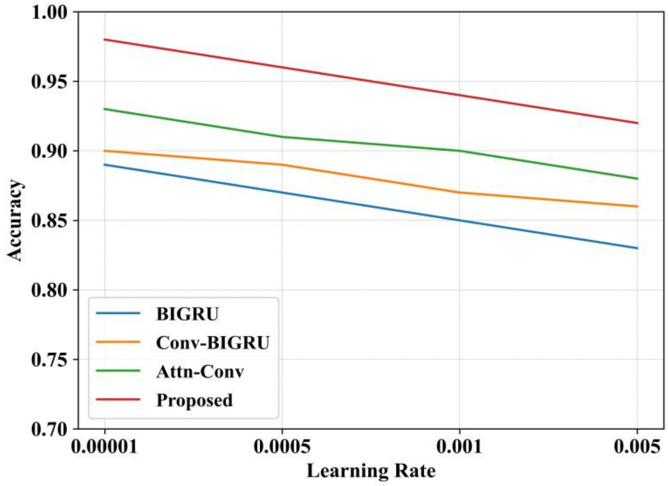

Figure 7: Accuracy comparison of proposed with existing works by varied learning rates.

Figure 7 shows the accuracy comparison of the suggested work with the present method when the learning rate is varied. When the learning rate is more than 0.001, the suggested method's accuracy is greater than 0.91%, BIGRU's accuracy is less than 0.85%, Conv-BIGRU's accuracy is less than 0.86%, and Attn-Conv's accuracy is less than 0.89%. By altering the epoch values, it is demonstrated that the suggested method has a greater accuracy value than the existing method.

Figure 8 depicts a comparison of the proposed method's accuracy with existing approaches by altering the batch size. When the batch size is 256, the suggested method's accuracy is better than 98%, BIGRU's accuracy is less than 93%, Conv-BIGRU's accuracy is less than 95%, and Attn-Conv's accuracy is less than 98%. As a result, it is demonstrated that the proposed approach's accuracy is worse when compared to the existing method by adjusting the batch size.

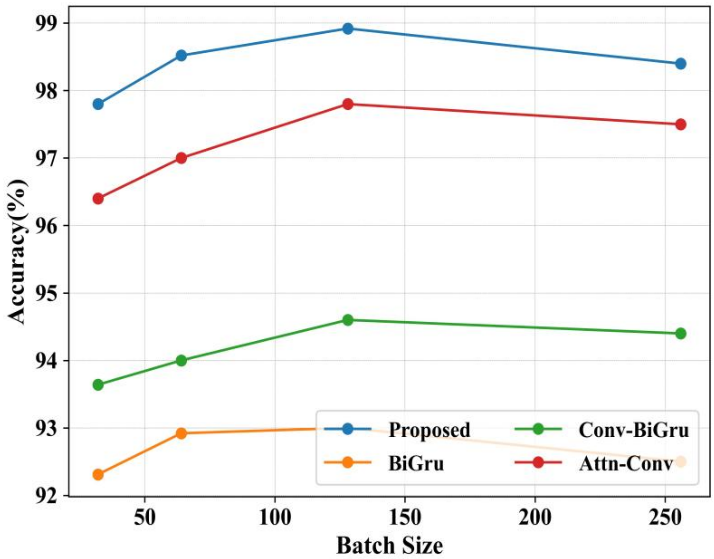

Figure 8: Accuracy comparison of proposed with existing worksby varied batch size.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

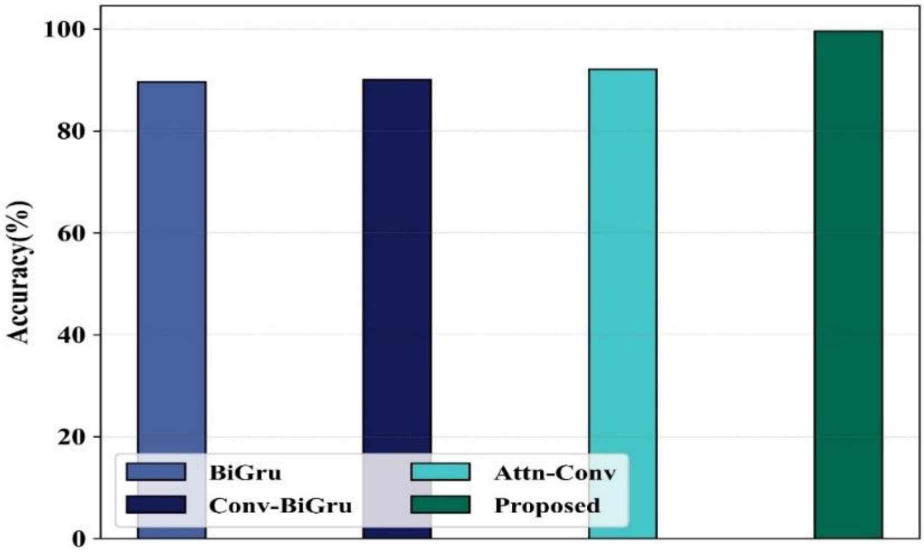

Figure 9: Accuracy comparison of proposed with existing works

Figure 9 compares the proposed method's accuracy to existing approaches such as BiGRU, Conv-BiGRU, and Attn-Conv. According to this graph, the proposed technique has an accuracy value of 0.996%, BiGRU has an accuracy value of 0.916%, Conv-BiGRU has an accuracy value of 0.898%, and Attn-Conv has an accuracy value of 0.877%. As a result, it is demonstrated that the proposed method has a greater accuracy value than the existing method. Figure 10 depicts a precision comparison between the proposed method and the existing methods.

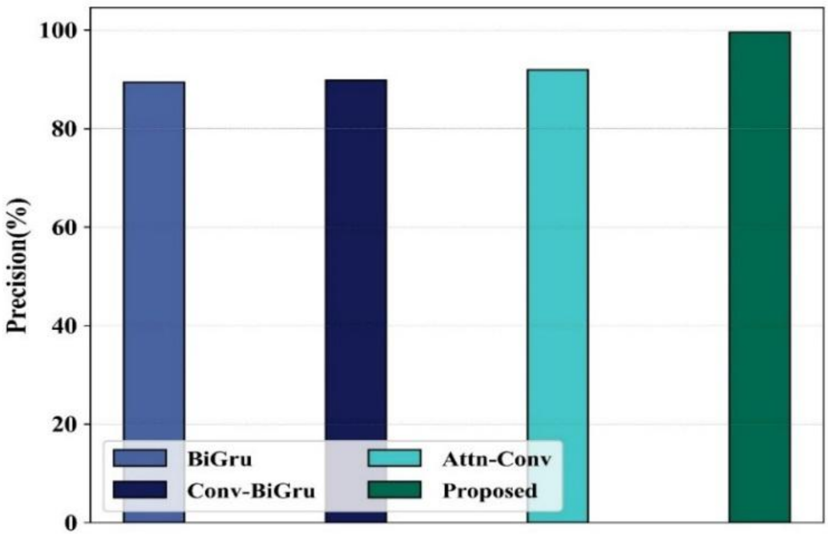

Figure 10: Precision comparison of proposed with existing works

Figure 10 shows how well the suggested method works compared to other methods like BiGru, Conv-BiGru, and Attn-Conv. The suggested method has a precision value of 0.996%, BiGru gets a precision value of 0.915%, Conv-BiGru gets a precision value of 0.896%, and Attn-Conv gets a precision value of 0.874%. Because of this, it has been shown that the proposed method is more accurate than the current way. Figure 11 shows how the proposed method compareswith existing methods in terms of recall.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

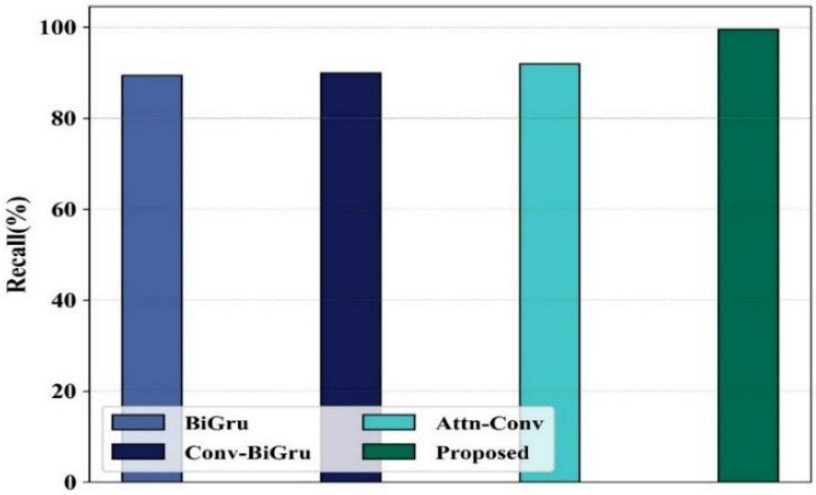

Figure 11: Recall comparison of proposed with existing works

Figure 11 shows the recall comparison of the proposed method with existing methods like BiGru, Conv-BiGru, and Attn-Conv. From this figure, it is derived that the recall value of the proposed method is 0.996%, BiGru achieves a recall value of 0.914%, the recall of Conv-BiGru is 0.894%, and Attn-Conv achieves a recall value of 0.875%. Thus, it is demonstrated that the proposed technique has a higher recall value than the existing method. Figure 12 compares the F1 score of the proposed technique to that of the existing methods.

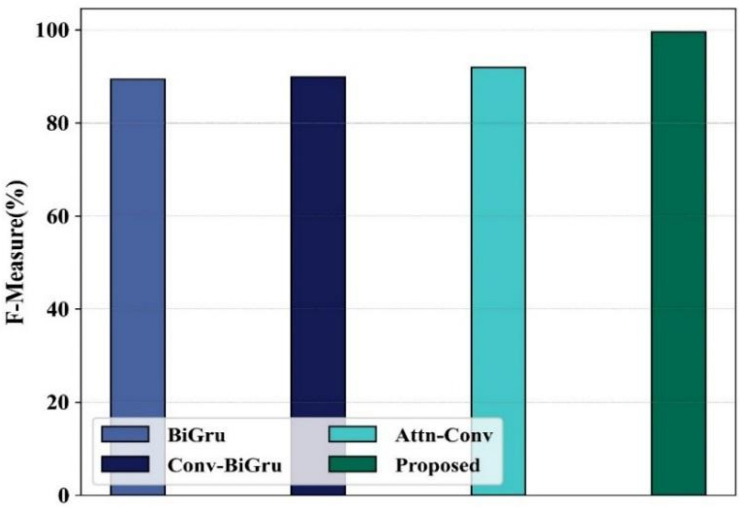

Figure 12: F1 Score comparison of proposed with existing works

The above figure signifies the F1 Score comparison of the proposed method with existing methods like BiGru, Conv-BiGru, and Attn-Conv. From this figure, it is inferred that the F1 Score value of the proposed method is 0.996%, BiGru attains an F1 Score value of 0.914%, the F1 Score of Conv-BiGru is 0.895%, and Attn-Conv gains 0.875% F1 Score. Therefore, it has been demonstrated that the suggested approach's F1 Score value is greater than the one used in the current method. The comparison of the proposed approach's Kappa coefficient with the existing methods is shown in Figure 13.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

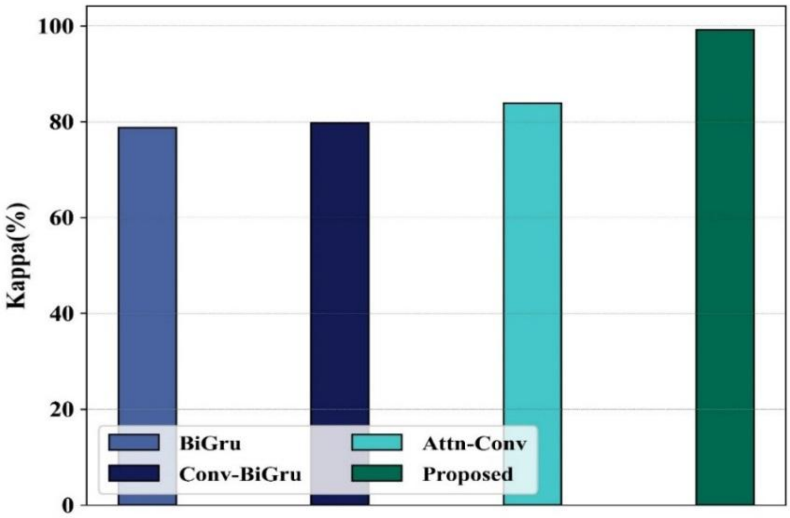

Figure 13: Kappa coefficient comparison of proposed with existing works

The proposed method's Kappa comparison with other approaches, such as BiGru, Conv-BiGru, and Attn-Conv, is shown in Figure 13. This figure indicates that the suggested method's Kappa value is 0.993%, BiGru achieves 0.829%, Conv-BiGru's Kappa is 0.791%, and Attn-Conv earns 0.750% Kappa. Therefore, it has been demonstrated that the suggested method's Kappa value is larger than the present method's. The MAE comparison between the proposed method and the existing methods is shown in Figure 14.

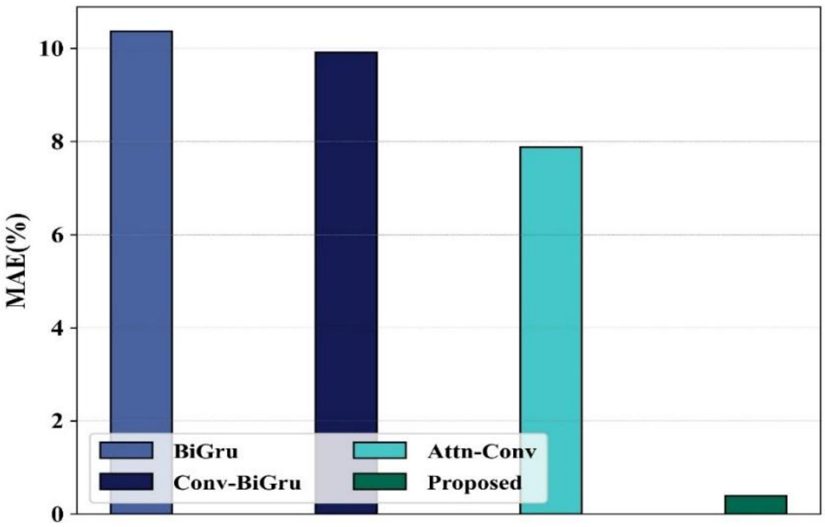

Figure 14: MAE comparison of proposed with existing works

The proposed method's MAE comparison with other approaches, such as BiGru, Conv-BiGru, and Attn-Conv, is displayed in the above figure. This statistic suggests that the MAE of the suggested approach is 0.003% that of BiGru is 0.083%, that of Conv-BiGru is 0.101%, and that of Attn-Conv is 0.122% MAE. Therefore, it is demonstrated that, in comparison to the current approach, the MAE value of the suggested method is small. The MSE comparison between the proposed and existing methods is shown in Figure 15.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

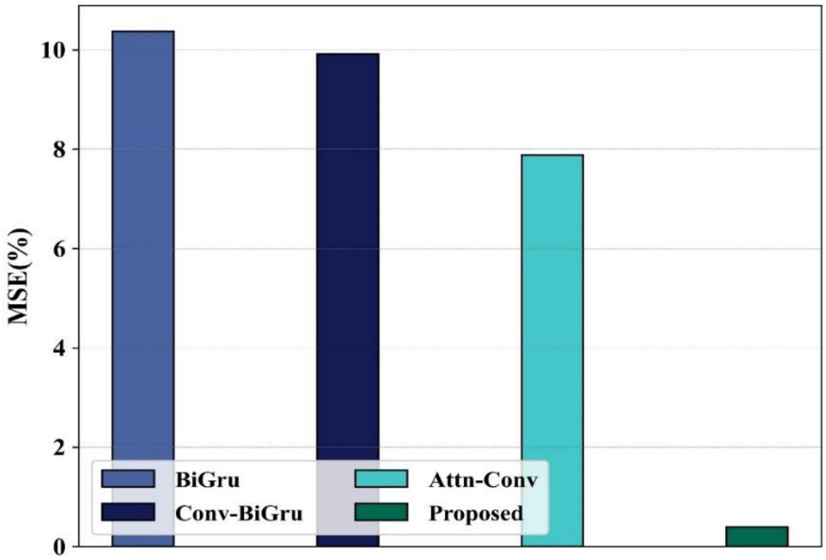

Figure 15: MSE comparison of proposed with existing works

The suggested method's MSE comparison with other approaches, such as BiGru, Conv-BiGru, and Attn-Conv, is shown in Figure 15. This figure suggests that the MSE of the suggested approach is 0.0032%, that of BiGru is 0.0833%, that of Conv-BiGru is 0.1017%, and that of Attn-Conv is 0.12% MSE. Therefore, it is demonstrated that, in comparison to the current method, the MSE value of the suggested method is negligible.

The suggested method's RMSE comparison with other approaches, such as BiGru, Conv-BiGru, and Attn-Conv, is shown in Figure 16. Based on this figure, can deduce that the suggested method's RMSE value is 0.057%, BiGru's RMSE value is 0.288%, Conv-BiGru's RMSE is 0.31%, and Attn-Conv gains 0.349% RMSE. Thus, it is demonstrated that, in comparison to the current method, the RMSE value of the suggested method is small.

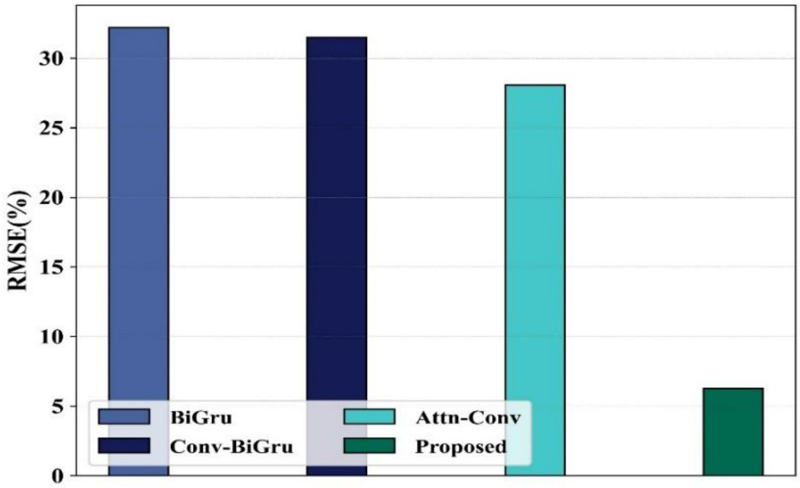

Figure 16: RMSE comparison of proposed with existing works

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

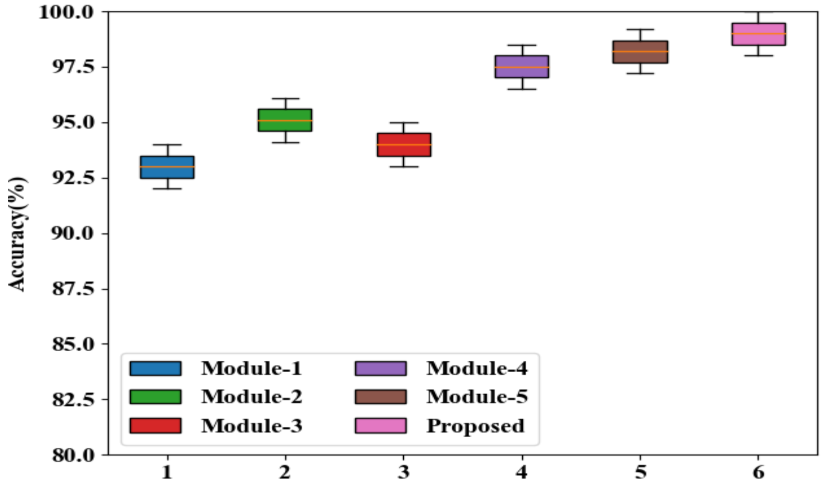

Figure 17: Ablation study of the proposed method by varied accuracy.

Figure 17 shows the ablation study of the proposed method with varied accuracy. The ablation was carried out in five distinct modules, namely modules 1, 2, 3, 4, and 5, as part of the proposed study. In this, module 1 represents the performance of the proposed method without using the pre-processing stage, module 2 is obtained by neglecting the weight parameters, module 3 exhibits the outcome of the proposed model by using only stemming and stop word removal, and module 4 illustrates that without using an optimization algorithm, module 5 demonstrated as without using feature extraction. The accuracy of module 1 is 93%, module 2 is 95.1%, module 3 is 94%, module 4 is 97.5% and module 5 is 98.2%.Thus, by comparing with all modules, the proposed study attains a better accuracy of 99%. This analysis proves the efficacy of each stage used in this proposed work.

- **Fake NewsNet dataset:** Figure 18 shows the comparison of training and testing accuracy arcs of the proposed method by varying the epoch values.

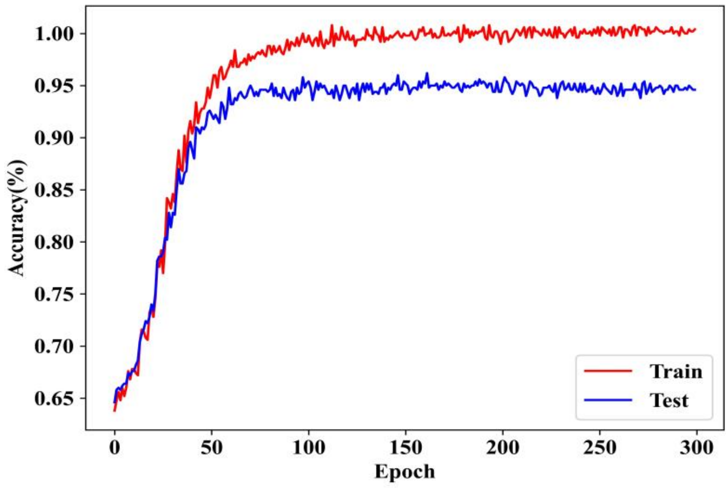

Figure 18: Training and testing accuracy

Figure shows the training and testing accuracy of the proposed method for changed epoch values. This figure shows that accuracy increases when the epoch value is close to zero and reaches the peak of training and testing accuracy of more than 0.91% when the epoch value is just above 50. After that, the accuracy curve remains stable values for the subsequent epoch. At higher epoch values, the proposed method shows higher accuracy for training and testing.

Figure 19 shows the training and testing loss of the proposed method for varying epoch values. This figure shows that the loss begins to decrease when the epoch value is just above zero and reaches the bottom of training and testing loss of less than 0.04% when the epoch value is just above 50. After that, the loss curve for success remains stable

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

epoch values. For higher epoch values, the proposed method shows minimal losses for training and testing.

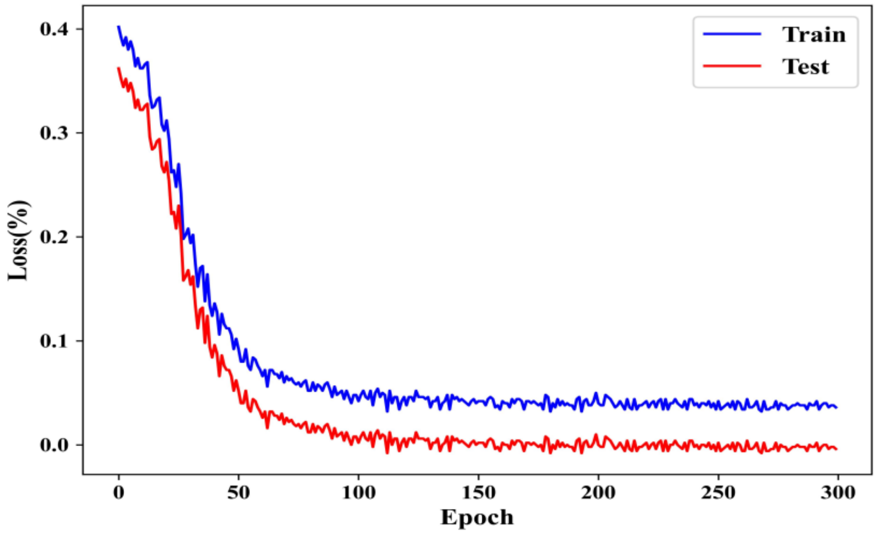

Figure 19: Training and testing loss

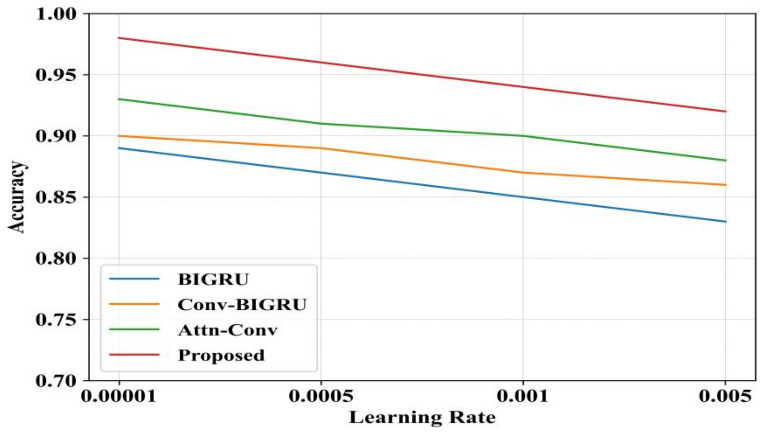

Figure 20: Accuracy comparison of proposed with existing works by varied learning rates.

By adjusting the learning rates, Figure 20 compares the accuracy of the proposed model with other existing methods by varying the learning rates. When the learning rate exceeds 0.001, the accuracy of the suggested approach is higher than 0.92%, BIGRU achieves an accuracy value lower than 0.85%, Conv-BIGRU achieves an accuracy value lower than 0.87%, and Attn-Conv achieves an accuracy value lower than 0.89%. Therefore, by adjusting the learning rate, it is demonstrated that the accuracy value of the suggested approach is higher than that of the current method. Figure 21 compares the accuracy of proposed and existing methods by varying batch sizes by altering the batch size.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

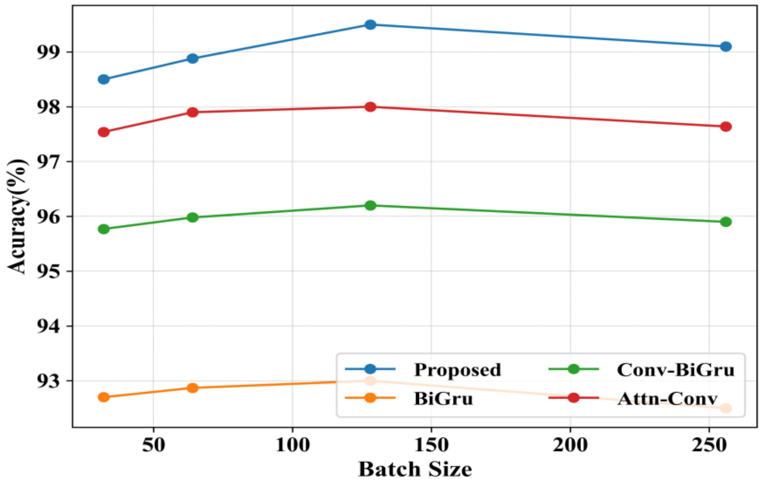

Figure 21: Accuracy comparison of proposed with existing works by varied batch size.

At 256 batches, the accuracy of the suggested technique is more than 99%, the accuracy of BIGRU is less than 93%, the accuracy of Conv-BIGRU is less than 97%, and the accuracy of Attn-Conv is less than 98%. Therefore, by adjusting the batch size, it is demonstrated that the accuracy value of the suggested approach is higher than that of the current method. The accuracy comparison between the suggested approach and the current methods is displayed in Figure 22.

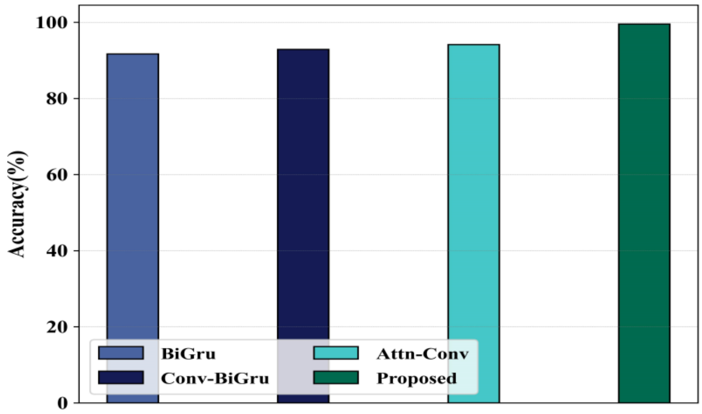

Figure 22: Accuracy comparison of proposed with existing works

The accuracy comparison between the proposed and existing methods,such as BiGru, Conv-BiGru, and Attn-Conv, is shown in Figure 22. This figure suggests that the accuracy value of the suggested approach is 0.997%, that of BiGru is 0.944%, that of Conv-BiGru is 0.925%, and that of Attn-Conv is 0.921%. Therefore, it has been demonstrated that the accuracy value of the suggested approach is more than that of the current method. The precision comparison between the proposed and existing methods is shown in Figure 23.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

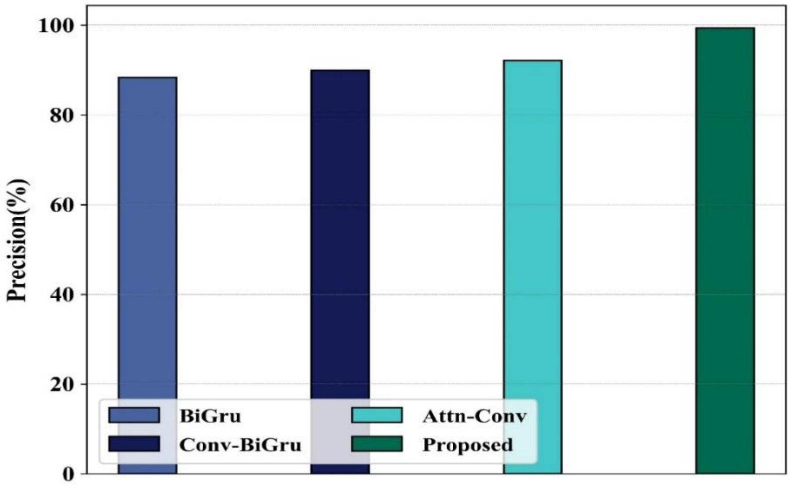

Figure 23: Precision comparison of proposed with existing works

This figure suggests that the precision value of the suggested technique is 0.997%, BiGru achieves 0.922%, Conv-BiGru achieves 0.897%, and Attn-Conv achieves 0.890% precision. Therefore, it has been demonstrated that the suggested method's precision value is higher than the existing method's.

Figure 24 compares the proposed method's recall to existing approaches such as BiGru, Conv-BiGru, and Attn-Conv. According to this graph, the proposed technique has a recall value of 0.994%, BiGru has a recall value of 0.923%, Conv-BiGru has a recall value of 0.894%, and Attn-Conv has a recall value of 0.888%. As a result, it is demonstrated that the proposed method has a higher recall value than the present method.

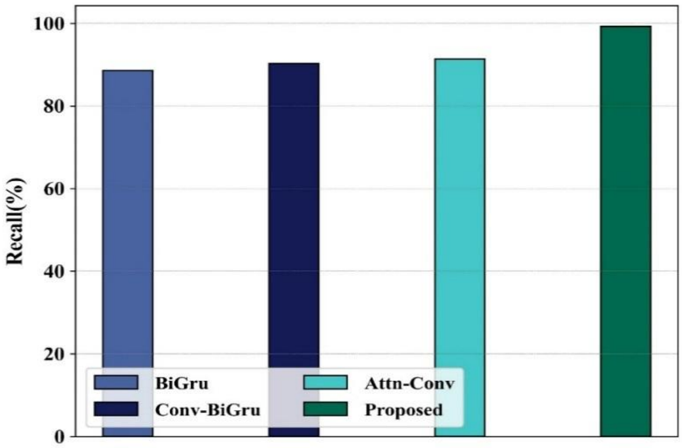

Figure 24: Recall comparison of proposed with existing works

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

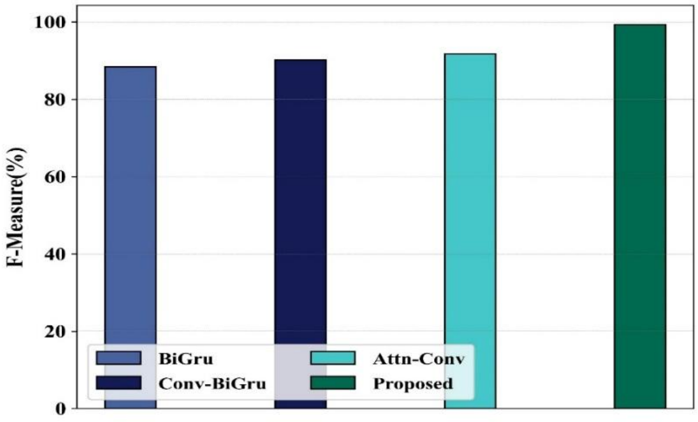

Figure 25: F1 Score comparison of proposed with existing works

Figure 25 depicts the suggested method's F1 Score comparison with existing approaches such as BiGru, Conv-BiGru, and Attn-Conv. According to this graph, the proposed approach has an F1 Score of 0.995%, BiGru has an F1 Score of 0.922%, Conv-BiGru has an F1 Score of 0.897%, and Attn-Conv has an F1 Score of 0.889%. As a result, it is demonstrated that the proposed method has a higher F1 Score value than the present method. Figure 26 depicts a comparison of the proposed approach's Kappa coefficient with the existing methods.

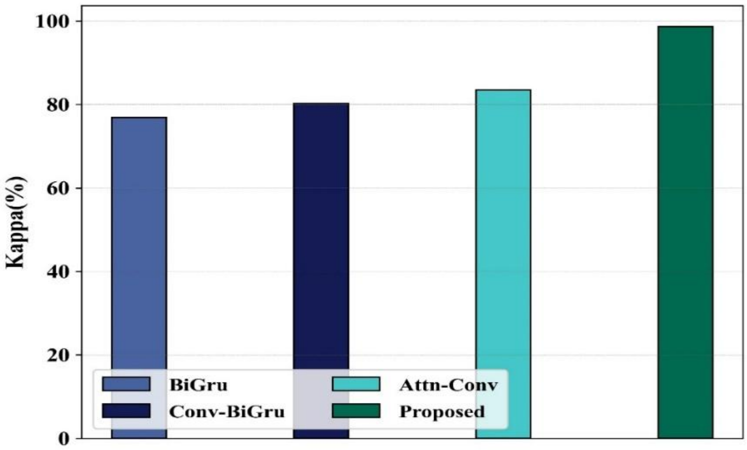

Figure 26: Kappa coefficient comparison of proposed with existing works

The Kappa comparison of the suggested method with known methods such as BiGru, Conv-BiGru, and Attn-Conv is shown in Figure 26 According to this figure, the proposed technique has a Kappa of 0.991%, BiGru has a Kappa of 0.845%, Conv-BiGru has a Kappa of 0.792%, and Attn-Conv has a Kappa of 0.779%. As a result, it is demonstrated that the proposed method has a greater Kappa value than the present method. The MAE comparison of the proposed method with the existing methods is depicted in Figure 27.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

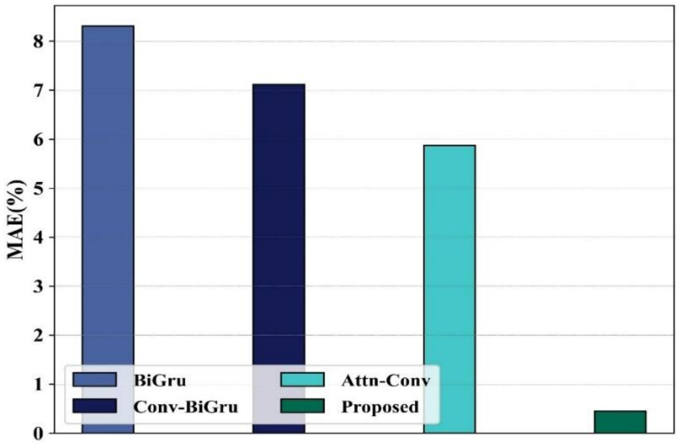

Figure 27: MAE comparison of proposed with existing works

The MAE comparison of the proposed method with known methods such as BiGru, Conv-BiGru, and Attn-Conv is shown in Figure 27. According to this graph, the proposed technique has an MAE of 0.002%, BiGru has an MAE of 0.055%, Conv-BiGru has an MAE of 0.074%, and Attn-Conv has an MAE of 0.078%. As a result, it is demonstrated that the proposed method has a lower MAE value than the present method. Figure 28 depicts the MSE comparison between the proposed approach and the existing method.

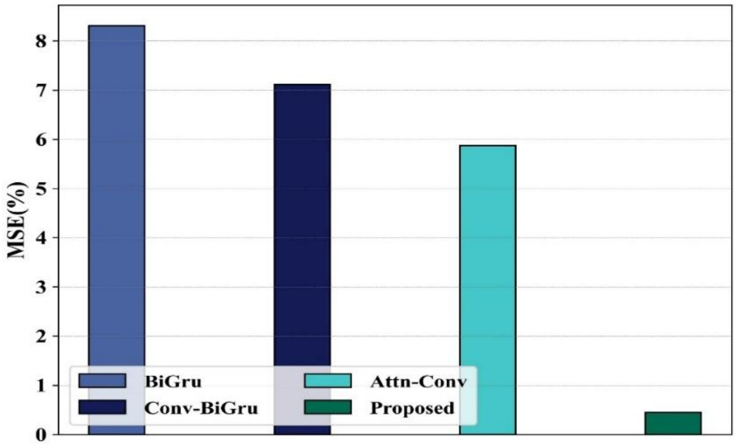

Figure 28: MSE comparison of proposed with existing works

The MSE comparison of the proposed method with known methods such as BiGru, Conv-BiGru, and Attn-Conv is shown in Figure 28. According to this figure, the proposed technique has an MSE of 0.0029%, BiGru has an MSE of 0.055%, Conv-BiGru has an MSE of 0.074%, and Attn-Conv has an MSE of 0.0788%. As a result, it is demonstrated that the proposed method has a lower MSE value than the present method. Figure 29 depicts the RMSE comparison between the proposed and existing methods.

Figure 29 compares the proposed method's RMSE to existing approaches such as BiGru, Conv-BiGru, and Attn-Conv. According to this figure, the proposed approach has an RMSE of 0.054%, BiGru has an RMSE of 0.235%, Conv-BiGru has an RMSE of 0.272%, and Attn-Conv has an RMSE of 0.280%. As a result, it is demonstrated that the proposed method has the lowest RMSE value when compared to the existing method.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

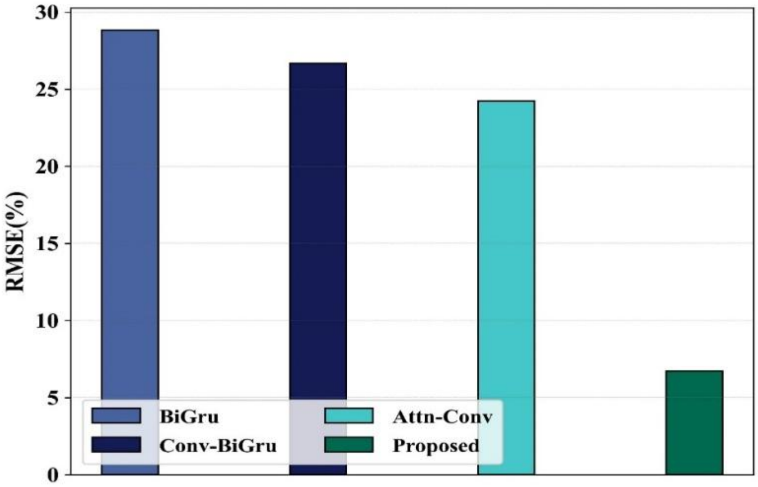

Figure 29: RMSE comparison of proposed with existing works

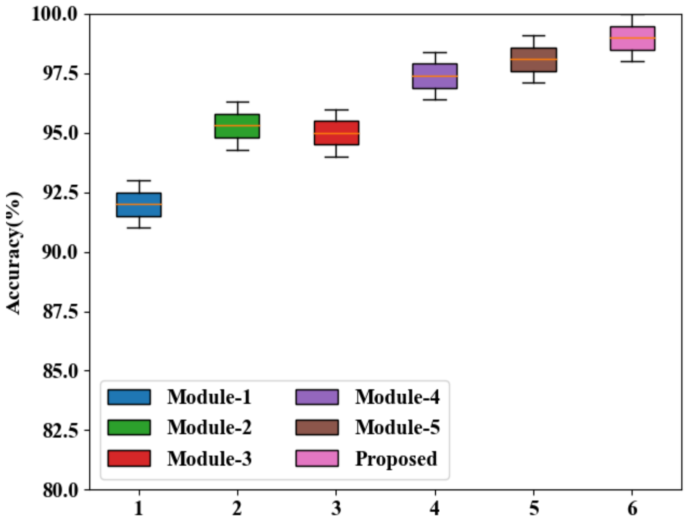

Figure 30: Ablation study of the proposed method by varied accuracy.

The ablation study analysis of the proposed work using the FakeNewsNet dataset is shown in Figure 30. By conducting an ablation study analysis, the strength of each stage that developed in this work is justified. Here, the ablation study is performed by five different modules: module 1, module 2, module 3, module 4 and module 5. Module 1 shows the accuracy results without the presence of pre-processing methods. Module 2 illustrates the results by neglecting the multi-attention mechanism. Module 3 shows the accuracy results without the utilization of the BiGRU network in the proposed model. Module 4 portrays the accuracy value without using NoR-BERT-based feature extraction method, and module 5 shows the accuracy value without fine-tuning the parameters. In the FakeNewsNet dataset, the attained accuracy of module 1 is 92%, module 2 is 95.3%, module 3 is 95%, module 4 is 97.4%, module 5 is 97.1%, and the proposed method achieved an accuracy of 99.1% in the ablation study. The inclusion of stages in the proposed work provides better accuracy. However, neglecting the use of pre-processing, feature extraction, multi-attention, BiGRU and fine-tuning process in the proposed study highly impacts the accuracy performance. Thus, the ablation results clearly justified the need for each stage developed in this study.

5. CONCLUSION

In this paper, a new Multi-attention assisted Convolutional Bi-GRU network is proposed for detecting fake news on the Twitter platform. Initially, pre-processing is done with the aid of effective NLP methods, which helps to reduce the unwanted irrelevant information. Then, the pre-processed data is given to the NoR_BeRT model to achieve best classification accuracy. Next, the optimal features are selected using the Du_BRO algorithm after going through the

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

feature dimensionality issues. The MA-TRC-BGRN model is utilized to detect fake news from the Twitter platform. Twitter EDA dataset provides lower optimization for MSE of 0.0032%, RMSE of 0.057%, and MAE of 0.003% and gains higher optimization for F1 Score of 0.9966%, accuracy of 0.99671%, and precision of 0.9967%. By usingthe fake news net dataset,the reduced error values are attained in terms of MSE of 0.029%, RMSE of 0.054%, and MAE of 0.002% and gains higher optimization for F1 Score of 0.995%, accuracy of 0.997%, and precision of 0.997%. The entire implementation of the work will be carried out in the Python platform. Finally, the model has been evaluated using the hybrid deep learning methodologies to prove its efficacy over the other existing models.However, the proposed study only detectsthe available fake news from Twitter platforms,and hence, this issue will be solved in the future by considering several social media platforms to further prove the efficiency of the proposed work. Also, the proposed study only considers text data for analyzing fake news so that complementary information is not attained, and this limitation will be resolved in the future by including image data along with the text to enhance the robustness of the proposed work. In future work, to gain a better outcome of estimation results, altered linguistics of social media platforms such as Greek, French, and Arabic will be considered. Also, in future studies, the work will be extended into real-time analysis.

REFERENCES

[1] Ali, A. M., Ghaleb, F. A., Al-Rimy, B. A. S., Alsolami, F. J., & Khan, A. I. (2022). Deep Ensemble Fake News Detection Model Using Sequential Deep Learning Technique. Sensors, 22(18), 6970.

[2] Alqurashi, Sarah, Btool Hamoui, Abdulaziz Alashaikh, Ahmad Alhindi, and Eisa Alanazi. "Eating garlic prevents COVID-19 infection: Detecting misinformation on the Arabic content of Twitter." arXiv preprint arXiv:2101.05626 (2021).

[3] Alyoubi, Shatha, Manal Kalkatawi, and Felwa Abukhodair. "The Detection of Fake News in Arabic Tweets Using Deep Learning." Applied Sciences 13, no. 14 (2023): 8209.

[4] Amoudi, Ghada, Rasha Albalawi, Fatimah Baothman, Amani Jamal, Hanan Alghamdi, and Areej Alhothali. "Arabic rumor detection: A comparative study." Alexandria Engineering Journal 61, no. 12 (2022): 12511-12523.

[5] Aslam, N., Ullah Khan, I., Alotaibi, F. S., Aldaej, L. A., & Aldubaikil, A. K. (2021). Fake detect: A deep learning ensemble model for fake news detection. Complexity, 2021, 1-8.

[6] Bahad, P., Saxena, P., & Kamal, R. (2019). Fake news detection using bi-directional LSTM-recurrent neural network. Procedia Computer Science, 165, 74-82.

[7] Capuano, N., Fenza, G., Loia, V., & Nota, F. D. (2023). Content Based Fake News Detection with machine and deep learning: a systematic review. Neurocomputing.

[8] Chauhan, T., & Palivela, H. (2021). Optimization and improvement of fake news detection using deep learning approaches for societal benefit. International Journal of Information Management Data Insights, 1(2), 100051.

[9] Choudhary, A., & Arora, A. (2021). Linguistic feature based learning model for fake news detection and classification. Expert Systems with Applications, 169, 114171.

[10] Djenouri, Y., Belhadi, A., Srivastava, G., & Lin, J. C. W. (2023). Advanced Pattern-Mining System for Fake News Analysis. IEEE Transactions on Computational Social Systems.

[11] Guo, Zhiwei, Qin Zhang, Feng Ding, Xiaogang Zhu, and Keping Yu. "A novel fake news detection model for context of mixed languages through multiscale transformer." IEEE Transactions on Computational Social Systems (2023).

[12] Hakak, S., Alazab, M., Khan, S., Gadekallu, T. R., Maddikunta, P. K. R., & Khan, W. Z. (2021). An ensemble machine learning approach through effective feature extraction to classify fake news. Future Generation Computer Systems, 117, 47-58.

[13] Hamdi, T., Slimi, H., Bounhas, I., & Slimani, Y. (2020). A hybrid approach for fake news detection in twitter based on user features and graph embedding. In Distributed Computing and Internet Technology: 16th International Conference, ICDCIT 2020, Bhubaneswar, India, January 9–12, 2020, Proceedings 16 (pp. 266-280). Springer International Publishing.

[14] Hiramath, C. K., & Deshpande, G. C. (2019, July). Fake news detection using deep learning techniques. In 2019 1st International Conference on Advances in Information Technology (ICAIT) (pp. 411-415). IEEE.

[15] Kaliyar, R. K., Goswami, A., & Narang, P. (2021). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimedia tools and applications, 80(8), 11765-11788.

[16] Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2020). FNDNet–a deep convolutional neural network for

fake news detection. Cognitive Systems Research, 61, 32-44.
[17] Konagala, V., & Bano, S. (2020). Fake News Detection Using Deep Learning: Supervised Fake News Detection Analysis in Social Media With Semantic Similarity Method. In Deep Learning Techniques and Optimization Strategies in Big Data Analytics (pp. 166-177). IGI Global.
[18] Kong, S. H., Tan, L. M., Gan, K. H., & Samsudin, N. H. (2020, April). Fake news detection using deep learning. In 2020 IEEE 10th symposium on computer applications & industrial electronics (ISCAIE) (pp. 102-107). IEEE.
[19] Kumar, S., Asthana, R., Upadhyay, S., Upreti, N., & Akbar, M. (2020). Fake news detection using deep learning models: A novel approach. Transactions on Emerging Telecommunications Technologies, 31(2), e3767.
[20] Luqman, Muhammad, Muhammad Faheem, Waheed Yousuf Ramay, Malik Khizar Saeed, and Majid Bashir Ahmad. "Utilizing Ensemble Learning for Detecting Multi-Modal Fake News." IEEE Access (2024).
[21] Luvembe, A. M., Li, W., Li, S., Liu, F., & Xu, G. (2023). Dual emotion based fake news detection: A deep attention-weight update approach. Information Processing & Management, 60(4), 103354.
[22] Monti, F., Frasca, F., Eynard, D., Mannion, D., & Bronstein, M. M. (2019). Fake news detection on social media using geometric deep learning. arXiv preprint arXiv:1902.06673.
[23] Mridha, M. F., Keya, A. J., Hamid, M. A., Monowar, M. M., & Rahman, M. S. (2021). A comprehensive review on fake news detection with deep learning. IEEE Access, 9, 156151-156170.
[24] Nadeem, M. I., Mohsan, S. A. H., Ahmed, K., Li, D., Zheng, Z., Shafiq, M., ... & Mostafa, S. M. (2023). HyproBert: A fake news detection model based on deep hypercontext. Symmetry, 15(2), 296.
[25] Nasir, J. A., Khan, O. S., & Varlamis, I. (2021). Fake news detection: A hybrid CNN-RNN based deep learning approach. International Journal of Information Management Data Insights, 1(1), 100007.
[26] Nawaz, A. N. A., Nawaz, M., Shaikh, N. A., Rajper, S., Baber, J., & Khalid, M. (2023). TPTS: Text Pre-processing Techniques for Sindhi Language: Text Pre-processing Techniques. Pakistan Journal of Emerging Science and Technologies (PJEST), 4(3).
[27] Ni, S., Li, J., & Kao, H. Y. (2021). MVAN: Multi-view attention networks for fake news detection on social media. IEEE Access, 9, 106907-106917.
[28] Rastogi, S., Gill, S. S., & Bansal, D. (2021, December). An Adaptive Approach for Fake News Detection in Social Media: Single vs Cross Domain. In 2021 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 1401-1405). IEEE.
[29] Saleh, Hager, Abdullah Alharbi, and Saeed Hamood Alsamhi. "OPCNN-FAKE: Optimized convolutional neural network for fake news detection." IEEE Access 9 (2021): 129471-129489.
[30] Sharma, A., Singh, I., & Rai, V. (2022, April). Fake News Detection on Social Media. In 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 803-807). IEEE.
[31] Swetha, P., & Priyanka, M. E. (2023). Fake News Detection on Social Media Using Regional Convolutional Neural Network Algorithm.
[32] Truică, Ciprian-Octavian, and Elena-Simona Apostol. "MisRoBÆRTa: Transformers versus misinformation." Mathematics 10, no. 4 (2022): 569.
[33] Umer, M., Imtiaz, Z., Ullah, S., Mehmood, A., Choi, G. S., & On, B. W. (2020). Fake news stance detection using deep learning architecture (CNN-LSTM). IEEE Access, 8, 156695-156706.
[34] Verma, Pawan Kumar, Prateek Agrawal, Vishu Madaan, and Radu Prodan. "MCred: multi-modal message credibility for fake news detection using BERT and CNN." Journal of Ambient Intelligence and Humanized Computing 14, no. 8 (2023): 10617-10629.
[35] Wani, A., Joshi, I., Khandve, S., Wagh, V., & Joshi, R. (2021). Evaluating deep learning approaches for covid19 fake news detection. In Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1 (pp. 153-163). Springer International Publishing.
[36] Xie, Bailin, and Qi Li. "Detecting fake news by RNN-based gatekeeping behavior model on social networks." Expert Systems with Applications (2023): 120716.
[37] Yildirim, G. (2023). A novel hybrid multi-thread metaheuristic approach for fake news detection in social media. Applied Intelligence, 53(9), 11182-11202.
[38] Zervopoulos, A., Alvanou, A. G., Bezas, K., Papamichail, A., Maragoudakis, M., & Kermanidis, K. (2020). Hong

Kong protests: using natural language processing for fake news detection on twitter. In Artificial Intelligence Applications and Innovations: 16th IFIP WG 12.5 International Conference, AIAI 2020, Neos Marmaras, Greece, June 5–7, 2020, Proceedings, Part II 16 (pp. 408-419). Springer International Publishing.

[39] Zhang, J., Dong, B., & Philip, S. Y. (2020, April). Fakedetector: Effective fake news detection with deep diffusive neural network. In 2020 IEEE 36th international conference on data engineering (ICDE) (pp. 1826-1829). IEEE.

[40] Zhang, Q., Guo, Z., Zhu, Y., Vijayakumar, P., Castiglione, A., & Gupta, B. B. (2023). A deep learning-based fast fake news detection model for cyber-physical social services. Pattern Recognition Letters, 168, 31-38.