IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 September 2024, accepted 2 October 2024, date of publication 9 October 2024, date of current version 11 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3476432

RESEARCH ARTICLE

A Hybrid Transformer-Based Model for Optimizing Fake News Detection

FATIMA AL-QUAYED [1], DANISH JAVED [2], N. Z JHANJHI [2], MAMOONA HUMAYUN [3], AND THANAA S. ALNUSAIRI [1]

[1]Department of Computer Science, College of Computer and Information Sciences, Jouf University, Sakaka 72311, Saudi Arabia
[2]School of Computer Science (SCS), Taylor's University, Subang Jaya 47500, Malaysia
[3]Department of Computing, School of Arts Humanities and Social Sciences, University of Roehampton, SW15 5PU London, U.K.

Corresponding author: N. Z Jhanjhi (noorzaman.jhanjhi@taylors.edu.my)

This work was funded by the Deanship of Graduate Studies and Scientific Research at Jouf University under grant No. (DGSSR-2023-02-02045).

ABSTRACT Fake news, colloquially referred to as false news, exerts a profound influence on fundamental facets of our societal framework. Manual fact verification is a frequently utilized strategy for mitigating the harmful impacts of false news transmission. However, when assessing the massive volume of freshly produced material, manual fact verification is insufficient. Furthermore, the quantity of labeled datasets is limited, people are unreliable annotators, resources are largely in the English language, and they mostly concentrate on articles related to news. Cutting-edge deep learning algorithms are employed to deal with this challenge automatically and address these difficulties. Nevertheless, the large number of models and variability of characteristics utilized in the literature frequently constitute a barrier for researchers attempting to enhance the effectiveness of models. This paper introduces a model designed to address the issue of false news, with a focus on analyzing news headlines. The approach is rooted in a hybrid classification model where our model integrates a BERT architecture, and the outputs are seamlessly linked to bi-directional deep learning layers, including a bi-LSTM layer and a bi-GRU layer. The model's training and evaluation were conducted using the WELFake dataset, comprising four prevalent news databases. A comparative analysis was carried out using the proposed model and standard classification models. Additionally, standard machine learning and deep learning models along with vanilla BERT were trained on the same dataset with comparable constraints to assess the impact of integrating a bi-directional deep learning layer with BERT. The results revealed an improvement in accuracy, as our proposed model achieved an accuracy of 98.1% and an F1 score of 0.982.

INDEX TERMS Bi-directional deep learning, transformers, fake news detection.

I. INTRODUCTION

In recent years, we have seen a surge in the phenomenon of false news. Fake news, otherwise known as false news [1], is thought to unsettle democracies, erode individuals' faith in public organizations, and have a significant impact on vital parts of our society, including elections, economic activity, and public perceptions. Several projects have begun to emerge to prevent the harmful impacts of the propagation of false news; one popular way to evaluate and dissect fake news is the verification of facts.

Manual verification internet sites like FactCheck.org use expert investigators to examine and discover false news; the goal of the veracity verifier is to authenticate established truths with information gathered from the news and judge the correctness of both. While manual verification of facts is important in identifying false news, it is insufficient for examining the massive amount of freshly produced content, particularly regarding fabricated information circulating on social media. Furthermore, the quantity of datasets labeled with false vs. regular news is inadequate, the accuracy of annotators lacks quality, and resources are largely in the English language, which focuses on just a subsection of news such as politics or sports, etc. [2]. Automatic false News

The associate editor coordinating the review of this manuscript and approving it for publication was Domenico Rosaci.

Detection algorithms have been created in recent years to address these challenges, and advanced models built with machine learning (ML) have been utilized to recognize false news and its harmful effects [3], [4], [5].

The objective of tackling the challenge of false news is to analyze it as a set of measurable indicators, usually, machine learning characteristics retrieved from articles. It can be an important instrument for determining the harmfulness of fake news; therefore, significant progress must be made by carefully selecting the best features and models. It is necessary to give researchers the material to increase the effectiveness of automated false news detectors to counteract the influence of misinformation on our society. Because of the range and complexity of models and characteristics utilized in the literature, research is typically sluggish and ineffective, causing uncertainty rather than aiding in developing stronger false news detectors. Fake news causes a lot of issues such as spreading rumors and lies to create chaos and confusion as shown in Figure 1.

FIGURE 1. Issues caused by fake news.

Machine learning (ML) and Deep Learning (DL) have received extensive attention in various areas such as social network analytics [6], sports analytics [7], sentiment analysis [8], autism detection [9], and other NLP (Natural Language Processing) tasks. However, many DL algorithms frequently face the problem of vanishing gradient, which limits the effectiveness of algorithms in learning from large databases, and this is the very limitation that Long Short-Term Memory (LSTM) attempts to solve. In the context of designing models for NLP tasks, word embedding is a critical consideration. To enhance this aspect, our proposed research leverages contextual word embeddings through the implementation of the BERT model [10].

BERT exhibits the capability to acquire contextualized word representations by leveraging extensive volumes of unlabeled text corpora [11]. Its impressive performance in various NLP tasks can be attributed to its intricate architecture and advanced nonlinear representation learning abilities. Bi-LSTMs and Bi-GRU play a crucial role in enhancing performance by effectively encoding and identifying essential information patterns. Consequently, the research harnesses the power of contextualized word representations from BERT, seamlessly integrating them with bi-directional DL to augment the performance of fake news classification. This integration capitalizes on the remarkable capacity of BERT to capture semantics and long-range dependencies within news headlines. Furthermore, there are several ways of talking about the challenge of fake news. Figure 2 below provides a classification of fake news attributes. The selection of attributes depends on the task and the available data. Since our data is text-based, therefore we will be focusing on text attributes.

One of the biggest limitations of existing fact-checking methods is their ability to generalize across multiple domains. Most existing models focus on the distribution of negative news but struggle to provide sufficient context or logic behind the prediction. While advanced models such as BERT have been applied to fake news detection with promising results, there remains a need for enhanced techniques that can more effectively capture the contextual and sequential dependencies in news articles which is critical in distinguishing between real and fraudulent content. We provide an extensive hybrid model for classifying false news that combines the BERT model with bi-directional deep learning to overcome these deficiencies. This method combines BERT's pre-trained language comprehension with the sequential pattern recognition capabilities of bi-directional deep learning layers (bi-LSTM, bi-GRU). The algorithm is intended to improve the accuracy of categorizing news stories as false or real, giving a strong response to the ongoing problem of disinformation across different domains.

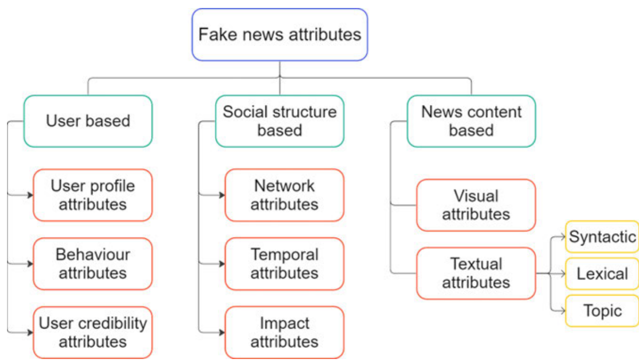

FIGURE 2. Fake news attribute classification.

Our research paper significantly enhances the field by providing novel insights into fake news as follows:

- Our research significantly advances the field by presenting a sophisticated model designed to classify fake news through the integration of the BERT model with bi-directional deep learning. This novel approach enables the accurate categorization of news articles as either fraudulent or authentic, thereby addressing the complex challenges posed by the proliferation of misinformation across multiple domains.
- This study rigorously investigates the efficacy of BERT+bi-LSTM and BERT+bi-GRU models, revealing notable performance enhancements in terms of accuracy and F1 score achieved through the integration of bi-directional deep learning layers with the BERT architecture.

- Our research demonstrates the model's efficacy in an English-language news setting by meticulously assessing and evaluating it against established categorization models and the baseline BERT framework. Our study contributes to a more nuanced understanding of the model's ability to distinguish between real and fake news stories by examining its performance over a wide range of linguistic nuances and news contexts.

This research is segmented into five sections. Section II is dedicated to the review of existing literature. Section III outlines the suggested models. Section IV contains information on implementation, which is followed by the assessment findings. Section V finishes the report by highlighting potential options for future research.

II. LITERATURE REVIEW

In this part, we provide the most impactful research within the area of fake news detection. The section concludes with a review of the current research. Table 1 provides the key literature related to fake news.

Several studies mentioned user profile-dependent characteristics for detecting false news. Reference [18] explored the connection between user profile traits and the authenticity of the news, concluding with new fake news detecting challenges. Reference [19] obtained an F1-score of 88% by using attention-oriented LSTM to categorize tweet content as rumor or otherwise using 13 languages. In another research [20], they created a support vector machine (SVM) framework for detecting false news based on three linguistic attribute categories, including writing structure, text com-

TABLE 1. Sentiment analysis key literature and datasets.

Cite	Aim	Challenges	Technique	Findings	Future work
[12]	To mitigate the possibility of propagation of false news.	Detecting false news, domain recognition, and bot detection in Twitter content.	Voting classifier and Bi-LSTM	Superior performances and high-impact features.	Employing attributes from veracity verification websites alongside Google searches to advance the identification of fake news.
[13]	To accurately evaluate the credibility of news.	Since just a single inference direction was used to categorize the viewpoint, some critical information may have been missed.	BERT language model	Improve the detection of the stance.	The suggested model's example of missed prediction is the stance relation of claim and article, which is in the "disagree" category.
[14]	To identify the fake news more precisely.	The needed information is frequently missing or insufficient at the initial stage, resulting in cross-domain interactions.	LSTM, depth LSTM, LIWC CNN, and N-gram CNN	Enhance the results, optimize the weights, and investigate the cross-domain interaction challenges.	The grammar evaluation would be thoroughly studied, and the subsequent processing would yield more important information, leading to increased precision in spotting false news.
[15]	Allow others to offer suggestions and help classify documents for evaluation.	Fake news identification at an early stage.	Neural Network based model	Offers great recall.	Contributing feedback can help fact-checkers determine if something is true or not.
[16]	To generate fake news pieces that involve both text and visual aspects, test error level analysis.	Social networking has become a major concern due to its capacity to have disastrous consequences by focusing on a specific type of news.	Hierarchical attention network, bidirectional GRU, Ensemble learning	When employing fake news instances, the greatest accuracy was 96%.	Fake news recognition via multimodal data is still a difficult and unexplored subject that requires additional research.
[17]	The efficiency of numerous ML approaches on 3 different datasets will be assessed.	Retaining modality-related properties affect the model's performance.	DL, ML, and pre-trained language models.	With small datasets, transformer-based algorithms, including BERT and others, excel in detecting fake news.	To recognize deceptive and health-related fake information on social networks amid the pandemic, such as COVID-19.

plexity, and psychology. Similarly, research [21] created linguistic characteristics from headlines and trained an SVM classifier manually. For news categorization, other investigators employed reinforcement learning [22] as well as fact-checking [23].

Reference [24] utilized the cosine similarity score to forecast credible content. They ran the suggested algorithm on one thousand stories and got an accuracy of 91.07% at a threshold of 0.62. In this research [25], they observed that generalizing language indicators for fake news identification across diverse topics and domains is a difficult task. Reference [26] devised a Z-value standardization strategy with 80-20% train and test split using the widely utilized linguistic characteristics software "Linguistic Analysis and Word Count (LIWC)." They examined many ML models and found that SVM produced the top outcomes with an accuracy of 87%.

In this research [27], authors suggested a BERT and CNN-oriented model while assessing the F1-score value in comparison to five existing cutting-edge models. They determined that their combination outperforms other similar works. They developed and evaluated their method using tweets based on three languages, claiming that it will perform better across other natural languages as well. Reference [28] designed and tested a one-layer CNN model and combined it with BERT. They attained an accuracy of 98.5%, although they emphasized that their method is only acceptable for smaller phrases and its resilience may be increased with several improvements.

Reference [29] developed a model for categorizing content into four distinct groups. They utilized Daily Mail stories as additional information for enhanced learning for CNN. They obtained an F1 value of 0.746. Authors of this research [30] introduced a hierarchical Bi-LSTM model and applied an attention mechanism for rumor identification. They were able to get an accuracy of 93.4% and 83.4% on two different databases. In related research [31], they employed a fine-tuned BERT on Korean language data to identify false news and obtained an AUC score of 83.8%.

Reference [32] used a deep learning-based technique to identify false news and reached an accuracy of 93.50%, proving that CNNs are effective for these sorts of tasks. In a different research [33], authors suggested a multifaceted deep learning-based technique for social network news categorization. For images, their research employed CNN, while RoBERTa was used for textual content, achieving an accuracy of 85.3% and 81.2% on two different datasets. Reference [34] described different existing tools and methods for detecting false news, as well as the involvement of veracity verification sites in this categorization process. Furthermore, they tested the LSTM and its bi-directional counterpart and determined that Bi-LSTM outperformed LSTM at an accuracy of 91.51%. Furthermore, the CNN processing layer takes a huge amount of information to train and is slower due to the max pool function. Additionally, there is an expectation of fully preprocessed and bigger data throughout the testing phase. Research [35] also employed a collective approach to assess the effectiveness of multiple ML-based methods, with AdaBoost and random forest achieving the greatest accuracy of 90.70%. Reference [36] introduced a unique method that relied on the occurrence of class labels and the distance vectorization technique for detecting false news and reported that their method had an accuracy of 97.5%.

Reference [37] detected bogus news with 98.36% accuracy using the GloVe and CNN. In related research [38], authors suggested another approach where BERT representations are supplied to the CNN algorithm for categorization, and it attained an accuracy of 98.90% using this technique.

Recent advances in NLP have seen an increased interest in transformer-based architectures such as BERT and GPT (Generative Pretrained Transformer) as they have achieved state-of-the-art results across various text classification tasks. BERT has been widely adopted in the area of fake news detection due to its ability to pre-train large-scale language data and fine-tune specific downstream tasks. Transformers employ a strong attention mechanism to analyze textual material continuously and construct complete and contextually aware word representations. A previous analysis found that BERT, a transformer-based model, outperformed the non-transformer method [39]. In this research [40], the authors propose a new hybrid fake news detection technique that incorporates BERT and a light gradient boosting machine (LightGBM) model. In our assessment of the literature, we discovered that propagation, linguistics, text semantics, and user profiles are relevant criteria for identifying false news. Furthermore, most of the research utilized ML and DL algorithms for false news identification that only considered the local context of textual data and ignored the global context. While models like BERT and its derivatives have achieved significant success in text classification tasks, their ability to incorporate temporal dependence is limited. Our model improves its ability to understand long-term dependencies by including bi-directional LSTM and GRU layers which are critical in discriminating between real and counterfeit news.

Transformer-based models like BERT and its derivatives have shown significant gains in managing linguistic subtleties and context-aware classifications within the field of fake news detection. These methods utilize self-attention processes to provide a solid platform for combating disinformation. Despite their success, all techniques have a limitation as they are not as effective at identifying sequential dependencies and temporal alterations in information which are critical for detecting increasingly complex types of fake news. Hybrid models attempt to address these issues through the trade-offs between accuracy and the capacity to predict sequence-based relationships, as well as the requirement for more generalizable models that perform well in a variety of news environments.

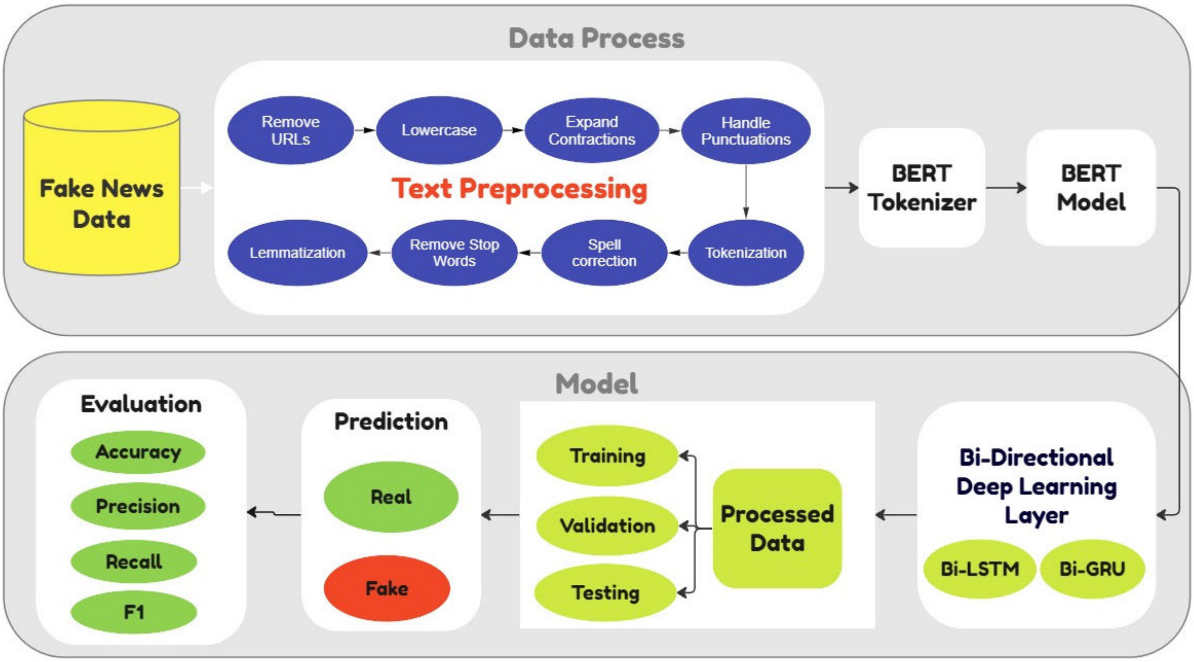

FIGURE 3. BERT-based Bi-directional deep learning model.

III. METHODOLOGY

This segment discusses the suggested model in more detail. We explain the conceptual model and method for the recognition of false news. In addition, our model is partitioned into multiple components to boost the results of fake news detection. Each module is well-detailed. Following this are guidelines for fine-tuning and effective utilization of the model. Figure 3 depicts our suggested model through visual representation, which consists of two phases.

A. PREPROCESSING

In the initial phase, our methodology initiates the implementation of a sequence of preprocessing procedures on the dataset. The preprocessing actions encompass tasks such as noise reduction and the elimination of redundant information. Fake news stories frequently contain irrelevant or misleading elements such as special characters, hyperlinks, HTML tags, and repetitive punctuation which do not provide useful information to the model. These elements can create excessive noise which confuses the model and impairs its ability to focus on key language patterns. We can make the input data cleaner and more typical of actual textual content by minimizing noise and allowing the model to focus on the attributes that are most important for identifying disinformation. This leads to greater generalization and increases the model's capacity to distinguish between true and fraudulent news based on important material rather than being distracted by certain characters or links. For example, deleting URLs guarantees that the model does not use hyperlinks to discriminate between genuine and fraudulent news as both categories frequently include links. The preprocessing actions encompass tasks such as noise reduction and the elimination of redundant information. The preprocessing phase is undeniably a pivotal component within the methodology of every Natural Language Processing (NLP) task [41], [42], [43]. The data cleaning process must be done before we supply the data to the transformer-based model. It includes handling punctuations which was done by employing a predefined Python function in conjunction with regular expressions to refine the text. Furthermore, we made an effort to eliminate alphanumeric characters, including hashtags and URLs. For each record, we transformed all uppercases into lower ones to reduce the size of the vocabulary, thus improving execution time. Moreover, we use spell correction to fix any spelling errors. Stop words are words in a phrase that contribute no meaning to it and whose removal has no effect on the text's processing for the intended purpose. They are deleted from the lexicon to minimize noise and the size of the feature set. Stopword removal enables the model to focus on content-rich phrases that give genuine cues to misinformation by removing high-frequency stopwords that do not distinguish between false and true news. Text normalization is the practice of simplifying many variants or tenses of a single word. It may be done in two ways: stemming and lemmatization. The former is the easiest of the two techniques. To stem a phrase, we just eliminate the suffix and reduce it to its root. Lemmatization is the method of reducing a word to its basic form. The main distinction between stemming and lemmatization is that lemmatization takes into account the context and transforms the word to its relevant base form while stemming just eliminates the final few letters, which sometimes results in wrong meanings and spelling problems. Lemmatization guarantees semantic consistency throughout the dataset by reducing each word to its lemma, which improves the model's ability to discern underlying sentence meaning. This is especially useful in false news detection as varied inflections or tenses can be utilized to misrepresent the facts. Thus, we apply lemmatization to change the words to their root form, thus reducing the size of the vocabulary for faster processing. Finally, the data is transformed into the refined format for the BERT transformer.

B. BERT

BERT includes an attention system designed to understand the context of content. BERT's architecture requires an encoder module (responsible for processing the text) and a decoder module (responsible for predicting the specific function). Transformer encoders process all messages simultaneously unlike sequential encoders which process encodings linearly which makes them non-directional. This allows the model to capture the meaning of each adjacent word, resulting in the bidirectional nature of BERT's basis for text classification. Input data must be prepared in the appropriate format prior to utilizing the pre-trained model. The model processes a list of token representations along with their associated attention masks for every encoder layer. The output consists of representations that are all the same size. The algorithm is provided with discrete vectors representing the completed sequence of instructions for classification purposes. Basically, the hidden state of the first tag [CLS] is used as the description of the whole sentence, thus making classification easier.

Fake news frequently manipulates language by employing unclear phrasing or recycling credible-sounding terms. BERT's capacity to build dynamic embeddings that capture the semantic meaning of words based on context aids in differentiating authentic news from fake news. BERT guarantees that linguistic subtleties are captured by learning deep semantic representations which allows the model to better recognize the minor differences between authentic and fraudulent information.

C. BERT INTEGRATION

Integration of BERT with Bi-GRU and Bi-LSTM to deal with false news can result in an effective model that benefits from both contextual embeddings and bidirectional sequence processing. We use a pre-trained BERT model to obtain contextual embeddings for each token in the input sequences. For each token, it stores a BERT embedding that captures the content of the document and the meaning of the word in the sentence. Then, we fine-tune BERT for fake news detection tasks to improve its performance. There are three main approaches to training the entire architecture to fine-tune a pre-trained BERT model on a smaller dataset as shown in Figure 4. First, we can retrain the whole pre-trained model, allowing all weights to be updated based on the new dataset. Second, we can freeze the initial layers of the pre-trained model and retrain only the higher layers, adjusting how many layers to freeze through trial and error. Third, the approach that we utilized is where we keep all pre-trained layers frozen and attach and train a few new neural network layers, updating only the weights of these new layers.

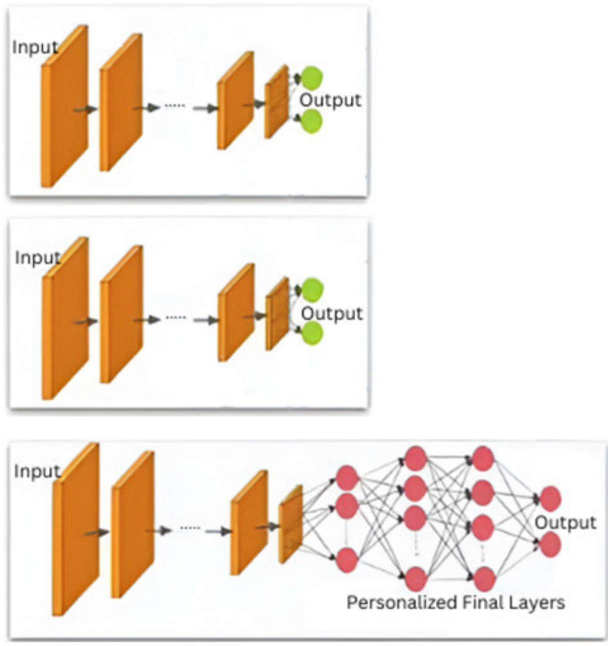

FIGURE 4. BERT Training- Top(Train entire architecture)- Middle (Train some while freezing others)- Bottom (Freeze the entire architecture).

The suggested approach employs the BERT-base-uncased modeling with a forward-feeding network containing hidden dimensions of 768. The BERT framework operates on input data represented by IDs and attention masks for sequences. BERT tokenizer is utilized to facilitate this task, taking input and adding [CLS] and [SEP] markers at the start and end. These processed inputs are subsequently supplied to the BERT framework. The BERT model generates representations for every token, each with a size of 768. These contextualized sentence-level representations, a hallmark of BERT, enhance the ability of Bi-LSTM and Bi-GRU to comprehend the semantics of the sentences. Recent research in this field has demonstrated that combining LSTM with word embedding models yields significant performance improvements [44]. Consequently, integrating bi-directional deep learning with BERT is expected to further enhance predictive accuracy, indicating a more profound understanding of semantic meaning within the proposed model. Leveraging BERT's bidirectional capabilities, a multi-layer encoding process is applied to the [CLS] token, capturing comprehensive information from each token and effectively acting as a "joint representation" well-suited for classification problems. The embeddings associated with the [CLS] token serve as the representation for the entire sentence and are provided as input to the classifier.

The classification module is constructed from the ground up and includes a forward-feeding layer where its value is set to 64. Batch normalization is employed to standardize inputs, and a dropout layer is set with a value of 0.2, which is introduced to avert overfitting. A dense layer of size 64 is set with the activation function 'relu'. Two additional forward-feeding layers are added where the output size is two; these output layers are incorporated to classify the news into two classes. A cut-off point is set to 0.8, which determines the output, favoring one of two classes.

Many fake news articles involve subtle alterations in meaning or misleading language that can only be understood if the sentence is read in context. Our model utilizes BERT to solve this issue by capturing word dependencies from both sides, ensuring that essential word associations are not neglected. This bidirectional processing enables our model to recognize complex language cues that may indicate disinformation, such as ambiguity, conflicting remarks, or exaggerated claims.

D. BI-DIRECTIONAL DEEP LEARNING

Bi-LSTM is a useful tool in detecting fake news. This is important for the model's ability to capture connections and nuances in data collection. Bi-LSTM provides two simultaneous accesses to different expressions and is good at understanding the meaning of words. This two-way process allows the model to understand not only the immediate content of the text but also its broader semantic content. Bi-LSTM is useful in detecting error patterns because fake news often contains subtle variations in terms and content. Its ability to memorize and model sequences of varying lengths allows it to detect disinformation that spans entire articles or news headlines. When combined with other techniques, Bi-LSTM improves the model's overall accuracy and efficacy in the essential task of detecting false news. The Bi-LSTM layer improves the model's capacity to monitor dependencies across longer lengths of text, allowing it to identify inconsistencies or patterns. This skill is critical for activities that involve comprehending how previous information connects to subsequent statements, such as spotting misleading claims hidden inside complicated narratives.

Bi-GRU's integration in the area of fake news identification can be beneficial as it provides an effective solution to the complex problem of detecting false information within news articles and headlines. Bi-GRU analyzes textual information in both forward and backward directions, allowing for a thorough comprehension of language's contextual complexities. This two-pronged approach allows the model to pick up on subtle social cues that fake news often uses. The GRU design improves the model's capacity to handle long sequences efficiently while minimizing the vanishing gradient problem with its gating features which is typically a difficulty in deep learning models. Bi-GRU's ability to discover patterns and correlations across extended textual content is invaluable in the setting of false news, where incorrect information may span numerous phrases or paragraphs. Bi-GRU is well-suited for fake news because it maintains Bi-LSTM's core capabilities such as capturing sequential dependencies while simultaneously decreasing computational cost. This is especially significant when dealing with dynamic or fast-moving news cycles because early detection of fake news is critical to limiting its spread.

IV. RESULTS

This section provides a comprehensive exposition of the results obtained from our experiment. It commences with a detailed description of the computer hardware and software configuration employed during the investigative process. Subsequently, various evaluation methodologies are discussed, highlighting the performance of our model relative to these benchmarks. The section delves into an exploration of distinct performance metrics, including accuracy, recall, and F1 measure. It is important to recognize that numerous factors, such as data pre-processing, exert a significant influence on fake news, particularly within specific contexts, as exemplified in our study concerning the WELFake dataset.

A. DATASET

In our research, we utilize the WELFake dataset, as introduced in [45]. This dataset is one of the largest available collections, comprising 72,134 news records, with 37,106 classified as fake and 35,028 as real. It adheres to the established criteria and guidelines for creating unbiased fake news datasets [46], which can be summarized as follows:

- All articles in the dataset must be expert-labeled.
- Fake news should be sourced from various origins.
- Genuine news should be obtained from reputable journalism organizations.
- A diverse range of news should be included to ensure a unique selection of trustworthy news.

The dataset includes the following attributes: news IDs, the title, and the text of the news. The news text and heading attributes had some indeterminate values at first. As a result, we integrated them and developed a new attribute to decrease indeterminate values while increasing the number of input sentences for better model learning. Furthermore, all records are categorized as follows: 0 is allocated to fake, while 1 is allocated to real news. The dataset is partitioned into two segments. The training dataset has been partitioned into 60% of the entire dataset, while the testing and validation set encompasses 20% each.

For dataset visualization, we use a histogram, which depicts the distribution of one or more variables by tallying the occurrences within distinct bins. Figure 5 shows the comparison of the word length between real and false news. False news seems to be quite a bit longer than real news, which is important information to consider when creating a model. Figure 6 provides a word cloud of fake news while Figure 7 shows the word cloud of real news. As we can see, some of the most important words are similar in both word clouds, which makes it challenging to distinguish between real and fake news. The word clouds for fake and real news show fascinating parallels and slight variances that highlight the difficulty of discriminating between them. In both clouds, we see that phrases like "said," "one," "people," "Trump," and "will" are similarly popular which indicates that fake and real news items frequently employ identical vocabulary, phrasing, and themes. As the surface-level linguistic characteristics are so similar, detection models may struggle to distinguish between real and fake news. That's why text preprocessing is one of the most important steps in fake news detection. The difficulty in discriminating between true and fraudulent news stems from similar lexical patterns which makes preprocessing techniques such as lemmatization or noise reduction critical. These serve to alleviate the confusion created by typical stop words or repeating sentences. Furthermore, the deep learning models utilized (e.g., BERT+bi-LSTM or bi-GRU) aid in processing these nuanced distinctions by leveraging context and recognizing language patterns at the sentence level instead of simply assessing individual word frequencies.

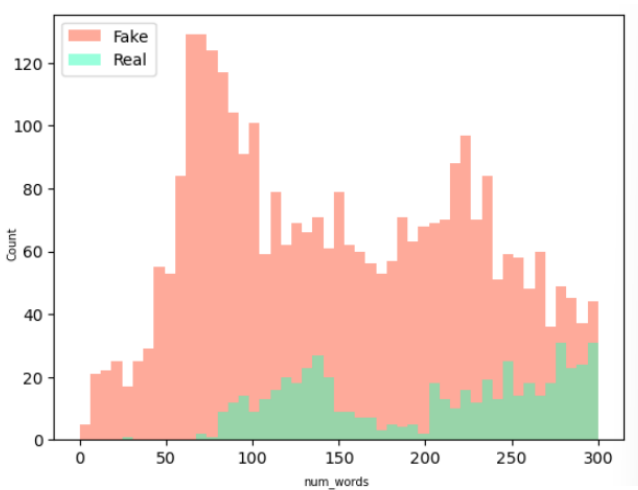

FIGURE 5. Distribution of news word count as real and fake.

FIGURE 6. Word cloud- fake news.

Most of the recent models employ one dataset, which fails to generalize on diverse datasets. Thus, we opt for a more comprehensive dataset, which is comprised of four different datasets. The WELFake dataset stands out for its expert-labeled articles as it follows recognized principles for dataset generation in fake news research. Expert labeling

FIGURE 7. Word cloud- real news.

guarantees that the dataset is not only accurate but also makes sure that it does not have any subjective biases that might emerge when using crowd-sourced or automated labeling approaches. WELFake recognizes the need to get fake news from a range of sources to reduce the danger of overfitting certain types or patterns of disinformation. The dataset enhances the generalizability of the model by adding false news records from multiple domains and sources which makes it more adaptive to detecting fake news from several platforms. Furthermore, the dataset contains actual news stories gathered from respectable organizations which ensures that the real news samples reflect trustworthy sources, thus providing comprehensive data for fake news detection.

B. EXPERIMENTAL SETUP

The trials were conducted by means of a machine equipped with Intel® Core™ i7-1185G7 Processor CPU, 24GB of DDR4 RAM, and a 1 TB SSD. The framework was designed and implemented using the Python language tool Spyder, which is a community-driven development platform for Python created by the firm Spyder project contributors.

C. EVALUATION METRICS

We construct four metrics that gauge the relationship for classifying among the two classes: true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN). Using these conditions, we assessed the results using various measures that are consistent with newer research in the area of ML [7] and DL [47].

Accuracy is the first measure we use for our study, which is defined in eq.1.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision is defined as the proportion of the number of accurate positive forecasts to the total number of positive forecasts.

$$\text{Precision} = \frac{TP}{FP + TP} \tag{2}$$

Recall is the proportion of the number of valid positive forecasts to the total number of accurately forecast outcomes.

$$\text{Recall} = \frac{TP}{FN + TP} \tag{3}$$

$$\mathrm{F1} = 2 * \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

D. CLASSIFICATION RESULTS

This section is dedicated to presenting the experimental outcomes, encompassing model training on the WELFake data and a comprehensive contrast of our suggested model against alternative methods. The learning process involved the random selection of 60% of the data points and 20% for the validation set, with subsequent testing conducted on the remaining 20%. The performance assessment of the models is underpinned by an array of evaluation criteria, encompassing metrics including accuracy and the F1 measure. These criteria serve as the basis for comparing the proposed model with its counterparts. The relevant comparative insights are visually conveyed through graphs, tables, and confusion matrices, offering a comprehensive evaluation of the suggested model. The proposed model combines BERT and bi-directional deep learning algorithms, including Bi-LSTM and Bi-GRU. The results of this implementation are subjected to a comparative analysis alongside several baseline models which are discussed in section V.

The findings reveal that the suggested model attains a remarkable accuracy of 98.1% and an F1 score of 0.982 for method 1, which combines BERT with Bi-LSTM. Similarly, our model provides an accuracy of 97.9% and an F1 score of 0.979 for method 2, which combines BERT with Bi-GRU, surpassing the performance of the other models, such as vanilla BERT, by 1.6% in accuracy. With regards to accuracy and F1 improvements over the baseline models, the suggested model exhibits a better performance based on experimentation on the WELFake dataset. The output scores of the suggested model are also compared with those of ML and DL algorithms to better comprehend the improvements. Table 2 displays the results below. Meanwhile, the visual representation of the outcome for the top-performing method 1 (Bi-LSTM + BERT) has been shown through the confusion matrix in Figure 8. Real news is represented by '0' while fake news is represented by '1'. The false positive rate (top right) is low (1.2%) which suggests that most real news articles are correctly identified. It is important as high false positives can lead to false alarms and potentially undermine trust in the detection system. Additional features such as source credibility analysis might assist the model in distinguishing between trustworthy news organizations and thus further mitigate this issue. The false negative rate (bottom-left) is 2.4% which means that these fake news articles are misclassified as real, and these undetected fake articles could contribute to the spread of misinformation. Incorporating claim-level verification tools like automated fact-checking APIs that cross-reference statements with credible sources can assist in eliminating false negatives. Furthermore, for method 2, the curves represent the accuracy and loss according to the number of epochs, as shown in Figures 9 and Figure 10. Similarly, for the best-performing method 1, the curves represent the accuracy and loss according to the number of epochs as shown in Figure 11 and Figure 12.

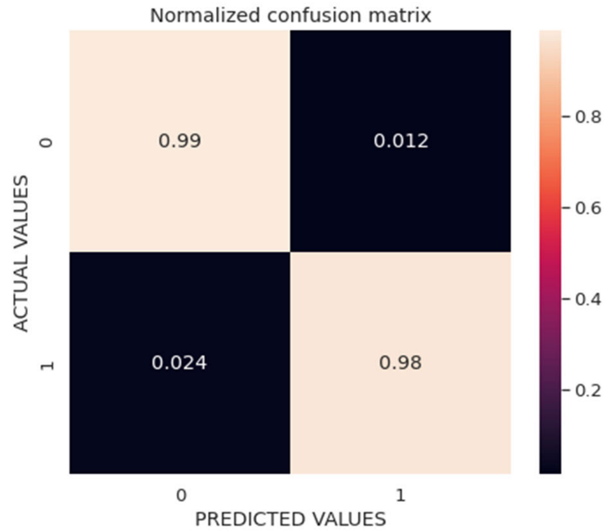

FIGURE 8. BERT+Bi-LSTM confusion matrix normalized.

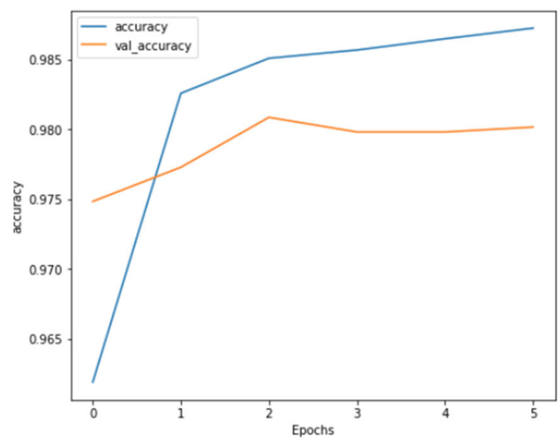

FIGURE 9. BERT+Bi-GRU validation accuracy.

TABLE 2. Accuracy report.

Algorithm	Validation Accuracy	Test Accuracy	F1 score (Test set)
Method 1: BERT + Bi-LSTM	98.4%	98.1%	0.982
Method 2: BERT + Bi-GRU	98.1%	97.8%	0.978
BERT	96.9%	96.5%	0.966
LSTM	96.5%	96.3%	0.96
Naïve Bayes	89.7%	89%	0.89
Random Forest	94.2%	93.7%	0.94

Our findings showed that the BERT+Bi-LSTM model performed somewhat better than the BERT+Bi-GRU model

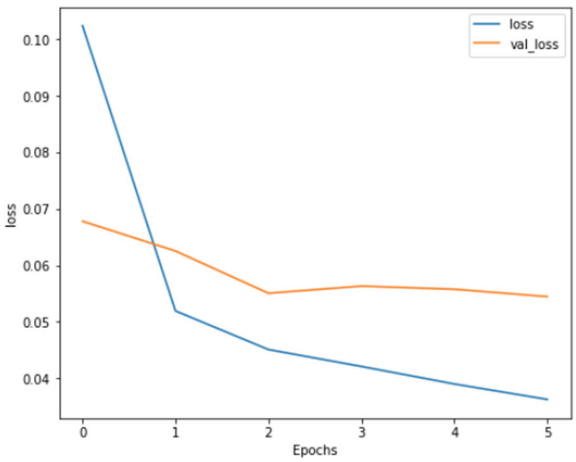

FIGURE 10. BERT+Bi-GRU validation loss.

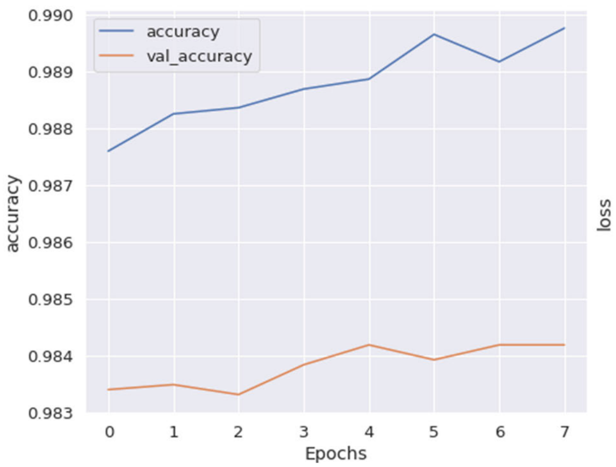

FIGURE 11. BERT+Bi-LSTM validation accuracy.

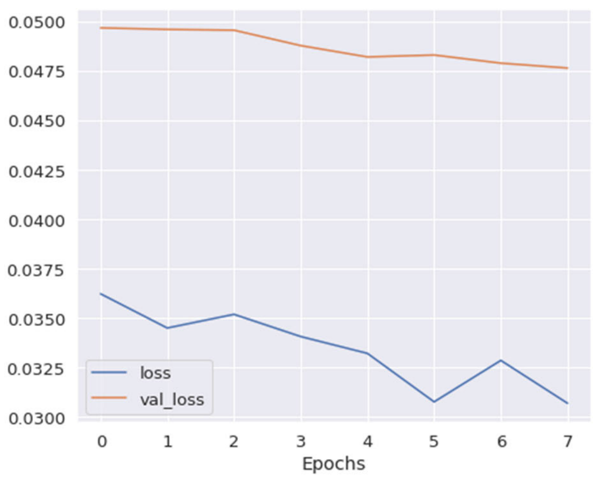

FIGURE 12. BERT+Bi-LSTM validation loss.

in terms of accuracy and F1 score. While the performance difference was not significant, it highlights the subtle variations in how Bi-LSTM and Bi-GRU designs handle sequential data and the trade-offs between complexity and computational efficiency. The slight performance advantage is due to Bi-LSTM's more robust handling of long-term dependencies and capacity to store more contextual information across sequences. While Bi-GRU improves computational efficiency and training time, it is slightly less capable of capturing the entire complexity of fake news, which needs nuanced language patterns and long-form connections. As a result of its higher parameter complexity and memory retention capabilities, Bi-LSTM achieved marginally superior accuracy and F1 score.

V. DISCUSSION AND COMPARISON

Deep neural networks have been demonstrated to be a valuable tool for the identification of fake news, and as such, we explored the efficacy of employing a combination of BERT with Bi-LSTM and Bi-GRU in our research. In our preliminary experiments, we examined the performance of various architectures, including standard machine learning and deep learning methods. These preliminary models yielded satisfactory results, but our quest for higher accuracy led us to explore more sophisticated approaches.

Our advanced models integrated BERT with Bi-LSTM and Bi-GRU into our architecture. BERT, a state-of-the-art language model, provides contextualized word representations, enhancing our model's understanding of semantics and context in news titles. Bi-LSTM, known for its ability to capture long-term dependencies in data, was used in conjunction with BERT to boost performance. The inclusion of these advanced components, along with attention mechanisms, significantly improved the performance of our models. We trained our data using an ensemble of BERT and Bi-LSTM, as well as an ensemble network that combined BERT and Bi-GRU, both equipped with attention mechanisms. Our research concluded that the BERT + Bi-LSTM ensemble model with an attention mechanism demonstrated the most promising results in fake news detection among the models we explored. Table 3 provides a comparison with other related studies. We test our model using two different methods. In method 1, we combine Bi-LSTM with BERT, and in method 2, we combine Bi-GRU with BERT. As we can see from Table 3, Bi-LSTM combined with BERT provides the best results.

Even though our proposed model performs well for the given data, real-life scenarios often involve more nuanced forms of fake news, such as inaccuracies in reports by otherwise reputable sources and the transformation of genuine news into fake news during its dissemination. While deep learning plays a crucial role in addressing these challenges, fact-checking for real-life events still depends on manual verification. Trusted veracity verification websites like PolitiFact and FactCheck.org remain essential in preventing news sources from being exploited. Additionally, addressing the issue of real news transitioning into fake news requires continuous monitoring of news stories shared by the public and comparing them with the original source.

If a semantic difference surpasses a predefined threshold, it may be classified as genuine news transformed into fake news.

TABLE 3. Result comparison.

Reference	Dataset	Algorithm	Accuracy
[45]	WELFake	ML algorithms	96.73%
[48]	WELFake	CNN+Bi-LSTM	97.74%
[49]	WELFake	LSTM+BERT	96.8%
	KaggleFakeNews		94%
[50]	ISOT	ML algorithms	96.36
Proposed Method 1	**WELFake**	**BERT+ Bi-LSTM**	**98.1%**
Proposed Method 2	WELFake	BERT+ Bi-GRU	97.8%

Our hybrid model provides several advantages over others in detecting fake news concerning accuracy and computational efficiency. However, two key areas where further research is necessary include handling multimodal data and the language dependency of the model. The current model is not equipped to analyze the visual components of fake news, therefore, it is less effective in detecting cases where images and text together contribute to the fake news content. Future iterations of the model could be extended to incorporate multimodal data processing by leveraging techniques such as visual-linguistic. Another challenge for fake news is language dependency, as our model has been primarily trained and evaluated on English-language data. While our model achieved great results in this scenario, its performance may not generalize as well to other languages without additional fine-tuning or adaptation. However, a little fine-tuning in the text pre-processing phase of the model can achieve high-quality results.

Both proposed model variations are computationally expensive because of the huge number of parameters in the BERT architecture and the additional complexity imposed by the bidirectional recurrent layers. The training requires significant computing resources including strong GPUs, a large memory bandwidth, and a significant amount of time for optimization. The bi-LSTM model has a larger computational cost than the bi-GRU model because it requires more resources due to its complex architecture, while the bi-GRU model offers a more efficient alternative with slightly reduced accuracy. For real-time fake news detection, models that can be trained and fine-tuned more efficiently would be advantageous. While the BERT+bi-LSTM model offers marginally better performance in terms of accuracy, the BERT+bi-GRU model's computational efficiency makes it more suitable for scenarios where speed and resource constraints are more critical than achieving the highest possible accuracy.

VI. CONCLUSION

Fake news, which is disseminated through mainstream channels like news outlets and social network platforms, often carries a deceptive agenda to influence public perception in a specific direction. The paper underscores the significance of fake news classification in today's information landscape and reviews the efforts made in this regard. Our research introduces a text-oriented model designed to categorize news articles as either genuine or fraudulent. To carry out this task, the model leverages BERT in conjunction with bi-directional deep learning as the classification framework. BERT exhibits the capacity to acquire contextualized word representations from an extensive corpus of unlabeled text, a feature that has contributed to its outstanding performance in numerous NLP applications owing to its robust architecture. Bi-LSTM and Bi-GRU, on the other hand, enhance the model's performance by effectively memorizing and recognizing critical information patterns. This was proven through experimentation on the WELFake dataset. Our proposed model achieved an accuracy of 98.1% F1 measure of 0.982.

The integration of contextualized word representations from BERT into a bi-directional deep learning layer is aimed at improving fake news classification, given their exceptional capability to grasp semantics and extended connections within news content. The suggested model is compared with alternative methods and a basic BERT framework, demonstrating a marginal enhancement. This enhancement signifies the model's ability to learn linguistic patterns within news headlines and their association with fake news. However, the model may face challenges in distinguishing fake news headlines from genuine ones, particularly when the language used by fake news authors closely resembles that of legitimate news sources. To address this issue, manual fact-checking of news headlines may be necessary. The paper highlights the potential for improving the model's performance with a larger dataset, enabling a more comprehensive analysis of language usage in fake news headlines.

Given the extensive spread of misinformation on social networks, future work could involve dealing with language patterns specific to social media-generated fake news. The proposed model holds promise for testing in various application domains, potentially enhancing existing benchmarks. Future investigations may involve tuning hyperparameters for both the BERT model and subsequent layers, accompanied by a thorough analysis of their impacts. We intend to expand our research in the future to involve graph-based techniques to detect fake news.

ACKNOWLEDGMENT

This work was funded by the Deanship of Graduate Studies and Scientific Research at Jouf University under grant No. (DGSSR-2023-02-02045).

REFERENCES

[1] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," *Inf. Process. Manage.*, vol. 57, no. 2, Mar. 2020, Art. no. 102025, doi: 10.1016/j.ipm.2019.03.004.

[2] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: A hybrid CNN-RNN based deep learning approach," *Int. J. Inf. Manage. Data Insights*, vol. 1, no. 1, Apr. 2021, Art. no. 100007, doi: 10.1016/j.jjimei.2020.100007.

[3] A. Abedalla, A. Al-Sadi, and M. Abdullah, "A closer look at fake news detection: A deep learning perspective," in *Proc. 3rd Int. Conf. Adv. Artif. Intell.*, 2019, pp. 24–28, doi: 10.1145/3369114.3369149.

[4] N. Aslam, I. U. Khan, F. S. Alotaibi, L. A. Aldaej, and A. K. Aldubaikil, "Fake detect: A deep learning ensemble model for fake news detection," *Complexity*, vol. 2021, no. 1, Jan. 2021, Art. no. 5557784, doi: 10.1155/2021/5557784.

[5] S. A. Alameri and M. Mohd, "Comparison of fake news detection using machine learning and deep learning techniques," in *Proc. 3rd Int. Cyber Resilience Conf. (CRC)*, Jan. 2021, pp. 1–6, doi: 10.1109/CRC50527.2021.9392458.

[6] D. Javed, N. Jhanjhi, N. A. Khan, S. K. Ray, A. A. Mazroa, F. Ashfaq, and S. R. Das, "Towards the future of bot detection: A comprehensive taxonomical review and challenges on Twitter/X," *Comput. Netw.*, vol. 254, Dec. 2024, Art. no. 110808, doi: 10.1016/j.comnet.2024.110808.

[7] D. Javed, N. Z. Jhanjhi, and N. A. Khan, "Football analytics for goal prediction to assess player performance," in *Proc. Int. Conf. Innov. Technol. Sports (RevealDNA ICITS)*, Apr. 2023, pp. 245–257, doi: 10.1007/978-981-99-0297-2_20.

[8] F. Long, K. Zhou, and W. Ou, "Sentiment analysis of text based on bidirectional LSTM with multi-head attention," *IEEE Access*, vol. 7, pp. 141960–141969, 2019, doi: 10.1109/ACCESS.2019.2942614.

[9] D. Javed, N. Zaman, N. A. Khan, D. G. W. Wei, and S. Balakrishnan, "Lily-Doll: Designing a system for the detection of autism spectrum disorder," in *Proc. 9th Int. Conf. Eng. Technol. Appl. Sci. (ICETAS)*, 2024, pp. 1–8.

[10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, vol. 1, Minneapolis, MN, USA: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.

[11] N. Liu and J. Zhao, "A BERT-based aspect-level sentiment analysis algorithm for cross-domain text," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–11, Jun. 2022, doi: 10.1155/2022/8726621.

[12] W. Antoun, F. Baly, R. Achour, A. Hussein, and H. Hajj, "State of the art models for fake news detection tasks," in *Proc. IEEE Int. Conf. Informat., IoT, Enabling Technol. (ICIoT)*, Feb. 2020, pp. 519–524.

[13] S.-X. Lin, B.-Y. Wu, T.-H. Chou, Y.-J. Lin, and H.-Y. Kao, "Bidirectional perspective with topic information for stance detection," in *Proc. Int. Conf. Pervasive Artif. Intell. (ICPAI)*, Dec. 2020, pp. 1–8, doi: 10.1109/ICPAI51961.2020.00009.

[14] Y.-F. Huang and P.-H. Chen, "Fake news detection using an ensemble learning model based on self-adaptive harmony search algorithms," *Expert Syst. Appl.*, vol. 159, Nov. 2020, Art. no. 113584, doi: 10.1016/j.eswa.2020.113584.

[15] Y. Yanagi, R. Orihara, Y. Sei, Y. Tahara, and A. Ohsuga, "Fake news detection with generated comments for news articles," in *Proc. IEEE 24th Int. Conf. Intell. Eng. Syst. (INES)*, Jul. 2020, pp. 85–90, doi: 10.1109/INES49302.2020.9147195.

[16] P. Meel and D. K. Vishwakarma, "HAN, image captioning, and forensics ensemble multimodal fake news detection," *Inf. Sci.*, vol. 567, pp. 23–41, Aug. 2021, doi: 10.1016/j.ins.2021.03.037.

[17] J. Y. Khan, M. T. I. Khondaker, S. Afroz, G. Uddin, and A. Iqbal, "A benchmark study of machine learning models for online fake news detection," *Mach. Learn. Appl.*, vol. 4, Jun. 2021, Art. no. 100032, doi: 10.1016/j.mlwa.2021.100032.

[18] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," 2017, *arXiv:1708.01967*.

[19] J. P. Singh, A. Kumar, N. P. Rana, and Y. K. Dwivedi, "Attention-based LSTM network for rumor veracity estimation of tweets," *Inf. Syst. Frontiers*, vol. 24, no. 2, pp. 459–474, Apr. 2022, doi: 10.1007/s10796-020-10040-5.

[20] B. D. Horne and S. Adali, "This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news," 2017, *arXiv:1703.09398*.

[21] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," in *Proc. 27th Int. Conf. Comput. Linguistics*. Santa Fe, NM, USA: ACL, Aug. 2018, pp. 3391–3401. [Online]. Available: https://aclanthology.org/C18-1287

[22] X. Zhou, A. Jain, V. V. Phoha, and R. Zafarani, "Fake news early detection: An interdisciplinary study," 2019, *arXiv:1904.11679*.

[23] N. Vo and K. Lee, "Learning from fact-checkers: Analysis and generation of fact-checking language," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.* New York, NY, USA: Association for Computing Machinery, 2019, pp. 335–344, doi: 10.1145/3331184.3331248.

[24] D. Mangal and D. K. Sharma, "Fake news detection with integration of embedded text cues and image features," in *Proc. 8th Int. Conf. Rel., Infocom Technol. Optim. (ICRITO)*, Jun. 2020, pp. 68–72.

[25] K. Sharma, F. Qian, H. Jiang, N. Ruchansky, M. Zhang, and Y. Liu, "Combating fake news: A survey on identification and mitigation techniques," 2019, *arXiv:1901.06437*.

[26] V. Singh, R. Dasgupta, D. Sonagra, K. Raman, and I. Ghosh, "Automated fake news detection using linguistic analysis and machine learning," in *Proc. Int. Conf. Social Comput., Behav.-Cultural Modeling, Predict. Behav. Represent. Modeling Simulation*, Jul. 2017, pp. 1–3, doi: 10.13140/RG.2.2.16825.67687.

[27] A. Safaya, M. Abdullatif, and D. Yuret, "KUISAIL at SemEval-2020 task 12: BERT-CNN for offensive speech identification in social media," in *Proc. 14th Workshop Semantic Eval.*, Dec. 2020, pp. 2054–2059, doi: 10.18653/v1/2020.semeval-1.271.

[28] C. He, S. Chen, S. Huang, J. Zhang, and X. Song, "Using convolutional neural network with BERT for intent determination," in *Proc. Int. Conf. Asian Lang. Process.*, Nov. 2019, pp. 65–70, doi: 10.1109/IALP48816.2019.9037668.

[29] H. Jwa, D. Oh, K. Park, J. Kang, and H. Lim, "exBAKE: Automatic fake news detection model based on bidirectional encoder representations from transformers (BERT)," *Appl. Sci.*, vol. 9, no. 19, p. 4062, Sep. 2019, doi: 10.3390/app9194062.

[30] H. Guo, J. Cao, Y. Zhang, J. Guo, and J. Li, "Rumor detection with hierarchical social attention network," in *Proc. Int. Conf. Inf. Knowl. Manage.*, 2018, pp. 943–951, doi: 10.1145/3269206.3271709.

[31] Y.-C. Ahn and C.-S. Jeong, "Natural language contents evaluation system for detecting fake news using deep learning," in *Proc. Int. Joint Conf. Comput. Sci. Softw. Eng.*, 2019, pp. 289–292, doi: 10.1109/JCSSE.2019.8864171.

[32] N. O'Brien, S. Latessa, G. Evangelopoulos, and X. Boix, "The language of fake news: Opening the black-box of deep learning based detectors," in *Proc. 32nd Conf. Neural Inf. Process. Syst. (NIPS)*, Montrèal, QC, Canada, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:54072129

[33] B. Singh and D. K. Sharma, "Predicting image credibility in fake news over social media using multi-modal approach," *Neural Comput. Appl.*, vol. 34, no. 24, pp. 21503–21517, Dec. 2022, doi: 10.1007/s00521-021-06086-4.

[34] D. K. Sharma, S. Garg, and P. Shrivastava, "Evaluation of tools and extension for fake news detection," in *Proc. Int. Conf. Innov. Practices Technol. Manage. (ICIPTM)*, Feb. 2021, pp. 227–232.

[35] M. Z. Khan and O. H. Alhazmi, "Study and analysis of unreliable news based on content acquired using ensemble learning (prevalence of fake news on social media)," *Int. J. Syst. Assurance Eng. Manage.*, vol. 11, no. S2, pp. 145–153, Jul. 2020, doi: 10.1007/s13198-020-01016-4.

[36] M. Mersinias, S. Afantenos, and G. Chalkiadakis, "CLFD: A novel vectorization technique and its application in fake news detection," in *Proc. 12th Lang. Resour. Eval. Conf.*, Marseille, France, May 2020, pp. 3475–3483. [Online]. Available: https://aclanthology.org/2020.lrec-1.427

[37] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "FNDNet—A deep convolutional neural network for fake news detection," *Cognit. Syst. Res.*, vol. 61, pp. 32–44, Jun. 2020, doi: 10.1016/j.cogsys.2019.12.005.

[38] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimedia Tools Appl.*, vol. 80, no. 8, pp. 11765–11788, Mar. 2021, doi: 10.1007/s11042-020-10183-2.

[39] S. F. N. Azizah, H. D. Cahyono, S. W. Sihwi, and W. Widiarto, "Performance analysis of transformer based models (BERT, ALBERT, and RoBERTa) in fake news detection," in *Proc. 6th Int. Conf. Inf. Commun. Technol. (ICOIACT)*, Nov. 2023, pp. 425–430, doi: 10.1109/ICOIACT59844.2023.10455849.

[40] E. Essa, K. Omar, and A. Alqahtani, "Fake news detection based on a hybrid BERT and LightGBM models," *Complex Intell. Syst.*, vol. 9, no. 6, pp. 6581–6592, Dec. 2023, doi: 10.1007/s40747-023-01098-0.

[41] S. N. Almuayqil, M. Humayun, N. Z. Jhanjhi, M. F. Almufareh, and D. Javed, "Framework for improved sentiment analysis via random minority oversampling for user tweet review classification," *Electronics*, vol. 11, no. 19, p. 3058, Sep. 2022, doi: 10.3390/electronics11193058.

[42] D. Javed, N. Z. Jhanjhi, and N. A. Khan, "Explainable Twitter bot detection model for limited features," in *Proc. Int. Conf. Green Energy, Comput. Intell. Technol. (GEn-CITy)*, Jul. 2023, pp. 476–481.

[43] D. Javed, N. Z. Jhanjhi, Q. M. Ijaz, A. K. Navid, and S. K. Ray, "Explainable machine learning based model for heart disease prediction," in *Proc. Int. Conf. Data Sci., AI Anal., Bridging Gap Between Practices (ICDSAIA)*, 2023, pp. 1–11.

[44] N. Rai, D. Kumar, N. Kaushik, C. Raj, and A. Ali, "Fake news classification using transformer based enhanced LSTM and BERT," *Int. J. Cognit. Comput. Eng.*, vol. 3, pp. 98–105, Jun. 2022, doi: 10.1016/j.ijcce.2022.03.003.

[45] P. K. Verma, P. Agrawal, I. Amorim, and R. Prodan, "WELFake: Word embedding over linguistic features for fake news detection," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 4, pp. 881–893, Aug. 2021, doi: 10.1109/TCSS.2021.3068519.

[46] G. Gravanis, A. Vakali, K. Diamantaras, and P. Karadais, "Behind the cues: A benchmarking study for fake news detection," *Expert Syst. Appl.*, vol. 128, pp. 201–213, Aug. 2019, doi: 10.1016/j.eswa.2019.03.036.

[47] M. Humayun, D. Javed, N. Jhanjhi, M. Fahaad Almufareh, and S. N. Almuayqil, "Deep learning based sentiment analysis of COVID-19 tweets via resampling and label analysis," *Comput. Syst. Sci. Eng.*, vol. 47, no. 1, pp. 575–591, 2023.

[48] M.-A. Ouassil, B. Cherradi, S. Hamida, M. Errami, O. E. Gannour, and A. Raihani, "A fake news detection system based on combination of word embedded techniques and hybrid deep learning model," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 10, pp. 525–534, 2022, doi: 10.14569/IJACSA.2022.0131061.

[49] N. Kausar, A. AliKhan, and M. Sattar, "Towards better representation learning using hybrid deep learning model for fake news detection," *Social Netw. Anal. Mining*, vol. 12, no. 1, p. 165, Dec. 2022, doi: 10.1007/s13278-022-00986-6.

[50] J. F. R. Bezerra, "Content-based fake news classification through modified voting ensemble," *J. Inf. Telecommun.*, vol. 5, no. 4, pp. 499–513, Aug. 2021, doi: 10.1080/24751839.2021.1963912.

FATIMA AL-QUAYED is currently an Assistant Professor with the College of Computer and Information Sciences, Jouf University, Saudi Arabia. She has multiple publications in WoS/ISI/SCI/Scopus. She has vast experience in academic qualifications. Her research interests include cyber security, wireless sensor networks (WSN), the Internet of Things (IoT), and knowledge management.

DANISH JAVED received the bachelor's and master's degrees in software engineering from Bahria University, Islamabad, Pakistan. He is currently pursuing the Ph.D. degree with the School of Computer Science, Taylor's University, Malaysia. His work is centered around utilizing advanced computational techniques to extract valuable insights from large datasets and enhancing the transparency and interpretability of AI models. He has contributed to various interdisciplinary projects aimed at enhancing the understanding and application of artificial intelligence in real-world scenarios. His work aims to advance the understanding of social networks and enhance the effectiveness of data-driven decision-making processes. His research interests include text mining, social network analytics, explainable AI, and big data analytics.

N. Z JHANJHI is currently a Renowned Scholar, a Professor, the Program Director of the Postgraduate Research Program in Computing, the Director of the Center for Smart Society 5.0 (CSS5), and the Cluster Head for Cybersecurity of the School of Computer Science, Taylor's University, Malaysia. His research portfolio spans more than 24 years, with a focus on cybersecurity, the IoT security, wireless networks, machine learning, and software engineering. He has an impressive track record of more than 700 research publications, including more than 400 SCIE high-indexed journal articles. His work has earned him an H-Index of 69 on Google Scholar, with more than 15,800 citations, underscoring his influence in the field. He has secured 13 international patents and has led 40 research grants as the Principal Investigator (PI) or the Co-PI. He is a sought-after keynote speaker, having delivered more than 60 invited talks globally. His contributions have solidified his reputation as a leading figure in his field, driving innovation, and excellence in academia and industry alike. Recognized globally, he has been listed among the World's Top 2% Scientists, in 2022 and 2023. He received the Vice Chancellor's Award for Most Cited Researcher at Taylor's University, in 2023. He serves as an editor for several high-impact journals. With extensive experience in academic quality accreditation and curriculum design, he continues to shape the future of cybersecurity and AI research, while mentoring the next generation of researchers through his supervision of M.Sc. and Ph.D. students.

MAMOONA HUMAYUN is currently a Senior Lecturer with the Department of Computing, School of Arts Humanities and Social Sciences, University of Roehampton, London, U.K. She has highly indexed publications in WoS/ISI/SCI/Scopus and her collective research impact factor is more than 200 points. Her Google Scholar H-index is 38 and i-10 index is more than 110, with more than 200 publications on her credit. She has several international patents on her account, including U.K. and Japanese. She edited/authored several research books published by world-class publishers. She has excellent experience in supervising and co-supervising postgraduate students and a good number of postgraduate scholars graduated under her supervision. She has completed several funded research grants successfully. She has served as a keynote/invited speaker for many international conferences and workshops. She has vast experience in academic qualifications, including ABET and NCAAA. Her research interests include cyber security, wireless sensor networks (WSN), the Internet of Things (IoT), requirement engineering, global software development, and knowledge management. She serves as a reviewer for several reputable journals.

THANAA S. ALNUSAIRI is currently a Lecturer with the College of Computer and Information Sciences, Jouf University, Saudi Arabia. She has multiple publications in WoS/ISI/SCI/Scopus. She has vast experience in academic qualifications. Her research interests include cloud computing, distributed systems, the Internet of Things (IoT), and P2P networks.