IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 1 February 2024, accepted 25 March 2024, date of publication 2 April 2024, date of current version 9 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3382832

RESEARCH ARTICLE

SA-Bi-LSTM: Self Attention With Bi-Directional LSTM-Based Intelligent Model for Accurate Fake News Detection to Ensured Information Integrity on Social Media Platforms

WANG JIAN[1], JIAN PING LI[2], MUHAMMAD ATIF AKBAR[3], AMIN UL HAQ[2], SHAKIR KHAN[4,5], REEMIAH MUNEER ALOTAIBI[4], AND SAAD ABDULLAH ALAJLAN[4]

[1]School of Artificial Intelligence, Neijiang Normal University, Neijiang, Sichuan 641100, China
[2]School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[3]Computer Science Department, Mohi-Ud-Din Islamic University, Nerian Sharif, Azad Kashmir 12010, Pakistan
[4]College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia
[5]University Centre for Research and Development, Department of Computer Science and Engineering, Chandigarh University, Mohali 140413, India

Corresponding authors: Jian Ping Li (jpli2222@uestc.edu.cn) and Amin Ul Haq (khan.amin50@yahoo.com)

This work was supported by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) under Grant IMSIU-RG23056.

ABSTRACT Detecting fake news is increasingly crucial in the digital age as social platforms and online news outlets amplify the spread of misinformation. This study proposes a novel approach for fake news detection using a Bi-LSTM neural network with an attention mechanism. The Bi-LSTM model's ability to discern long-range relationships and sequential patterns in text data is augmented by the attention mechanism, focusing on key segments within the text to enhance discriminative power. The dataset for training comprises a diverse collection of news articles, meticulously annotated by human fact-checkers to include both real and fake news samples. Pre-processing steps, including tokenization, are undertaken to optimize the model's learning capabilities. The proposed Self-Attention with Bi-Directionals-LSTM (SA-BiLSTM) model demonstrates promising performance in distinguishing real from fake news. The attention mechanism allows the model to highlight critical words and phrases, capturing essential features for accurate predictions. Evaluation through cross-validation reveals competitive results as opposed to cutting-edge fake news identification methods. Ablation studies emphasize the attention mechanism's significant contribution to identifying relevant textual patterns associated with fake news. The model, incorporating CNN, GRU, Attention-GRU, LSTM, and Bi-LSTM, surpasses contemporary procedures, as validated using the ISOT dataset. The results demonstrate a robust approach to fake news detection using Bi-LSTM with attention, contributing to the fight against misinformation. The proposed model SA-Bi LSTM obtained 99.98% predictive accuracy on test data. The accuracy of the proposed model (SA-BiLSTM) comparatively to baseline models is higher. The model shows potential for preserving information integrity on social media and news outlets, bolstering public trust in reliable sources.

INDEX TERMS Fake news detection, long short-term memory (LSTM), bid-LSTM, self-attention mechanism, SA-BiLSTM, ISOT dataset.

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

I. INTRODUCTION

In recent times, the rapid proliferation of digital platforms and the widespread dissemination of information have presented a

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

formidable challenge: differentiating between accurate news and misleading or fabricated articles [1]. The surge of fake news characterized by the deliberate spread of misinformation or misleading content masquerading as genuine news, has highlighted the critical need for efficient and automated techniques to identify and resolve this problem [2].

The advancement of deep learning algorithms has opened up new solutions for the challenge of identifying bogus news. In natural language processing, deep learning approaches including transformer models like BERT, recurrent neural networks (RNNs), and convolutional neural networks (CNNs) have shown promise in two applications: text categorization and sentiment analysis [3].

Social media now permeates every aspect of our culture, significantly influencing our daily lives, routines, and habits. It has revolutionized information distribution and consumption, with messaging, blogging, and social micro-journalists transforming the way we access news compared to traditional media distributors. Users actively engage in sharing and commenting on news articles, leading to a rapid information diffusion process within social networks. Social media's growing popularity has caused consumers to change the way they obtain news, with web-based news becoming the primary source of data. Nevertheless, a significant amount of information found on the Internet is questionable and may even be deliberately misleading [4]. The impact of false news has been especially evident in significant events such as the 2016 U.S.A. presidential election campaign. Many believe that false news played a substantial function in influencing public consultation during the election. Consequently, the term "fake news" has gained prominence in mainstream conversations and has become an integral part of our modern vernacular [5].

In the present day, the proliferation of false news has become a significant worry for both academic and industrial. One of the suggested remedies to combat this issue is rectified by humans. However, the ever-changing and instantaneous nature of fake news on social media makes it much more difficult to distinguish false information online [5].

Expert fact-checking may offer limited assistance due to its low efficiency. Furthermore, relying solely on human fact-checking is laborious and costly. In light of these challenges, utilizing Deep Learning Models (DL) and Machine Learning (ML) becomes imperative to motorize and expedite the procedure. Using hierarchical classification techniques, bogus news may be identified. By leveraging DL techniques, We can improve false news detection's precision and efficacy, reducing the need for human fact-checkers and enabling a more efficient and scalable approach to combat misinformation [6].

From the above discussion and from the literature review we reached the conclusion that existing fake news detection models do not effectively detect fake news. To tackle the accurate detection problem of false news, a new architecture is necessary for accurate detection of false news. Plenty of work is done by researchers related to Fake News Detection but is not abundant. Based on the previous work, different researchers use different algorithms, but there are still some limitations like time, cost, accuracy, etc. In order to tackle the problem of lack of prediction accuracy, we proposed a new model using deep learning techniques in this work. In designing the model, we will incorporate deep learning techniques such as data pre-processing, attention mechanism, and LSTM and Bi-LSTM. The proposed method will evaluated using Fake new dataset ISOT.

In this study, we have developed an improved Bi-LSTM model for false news identification that is based on artificial intelligence (AI). In the designing of the suggested model self-attention approach will be merged with Bi-LSTM to boost the predictive potential of the model. Then the proposed model Self-Attention +BiLSTM (SA-BiLSTM) will be validated using a Fake news data set. The suggested mode will be trained and tested using the held-out cross-validation approach. The performance assessment metrics will be computed in order to measure the model's effectiveness. Additionally, we will contrast the suggested model's prediction ability with that of the most recent models. The major contributions of this study as follows:

1) For precise false news identification, the Self-Attention+BiLSTM (SA-BiLSTM) AI-based model is suggested.
2) LSTM, Bi-LSTM, and Bi-LSTM with attention mechanism models will be evaluated using fake news processed data.
3) The recommended model's performances were evaluated using a range of performance assessment metrics.
4) When compared to baseline models, the suggested model performs well.

The remaining portions of the paper are organized in this manner. The paper's subsequent sections are arranged so that the literature review is provided in section II. The specifics of the data sets and classification algorithms utilized in this work were covered in section III. Section IV presents the outcomes of the experiment. Section V has covered the conclusion and the direction of future research.

II. LITERATURE REVIEW

To identify and locate false information, researchers have developed a wide range of machine learning and deep learning techniques. In this research study, we have focused on giving some basic techniques for spotting fake news. The primary goal of our literature study is to highlight the drawbacks of these conventional methods and offer a reliable solution. In an attempt to identify the top-performing machine learning model, several academics have conducted extensive analyses of a wide variety of models on a variety of datasets.

Five machine learning (ML) and three deep learning (DL) models have been applied to the ISOT and KDnugget datasets. These models include classification algorithms such as support vector machine (SVM), decision tree (DT), k nearest neighbor (k-NN), random forest (RF), logistic regression (LR), and advanced architectures like long short-term memory (LSTM), convolutional neural networks (CNN), and gated recurrent networks (GRU). The goal of these models is to identify bogus news material. In light of the outcomes of their experiments, the Random Forest algorithm demonstrated superior performance on ISOT Dataset and Logistic Regression on KDnugget Dataset compared to other intelligent classification algorithms [4].

The 19 models assessed fall into three categories: 8 traditional learning (TL), 6 deep learning (DL), and 5 pre-trained language models BERT (Bidirectional encoder representations from transformers). BERT-based models consistently exceed all other models across all datasets. The fact that the BERT-based models with previous training demonstrated resilience even with small dataset sizes and performed noticeably better than other models is very remarkable [7].

The ensemble classification methodology put forth for identifying false information has demonstrated superior accuracy when in comparison with existing cutting-edge techniques. In this novel model, significant attributes are extracted from datasets containing fabricated information. After that, an ensemble model made up of three well-known machine learning algorithms—Random forest (RF), decision tree (DT), and extra tree classifier—is used to classify these characteristics (ETC). Remarkably, on ISOT their prototype has achieved noteworthy training 99.8% and testing 44.15% accuracy. Similarly, when applied to the Liar dataset, our model has attained flawless training and testing accuracies of 100%. These outcomes underscore the efficacy and resilience of our proposed ensemble model in precisely discerning instances of bogus news [8].

An innovative hybrid deep learning framework, integrating RNN and convolutional, has been innovatively designed for the explicit purpose of categorizing fabricated news. This hybrid model has been meticulously validated using two distinct fake news datasets: ISOT and FA-KES. The outcome of the validation procedure conclusively illustrates that the suggested hybrid model surpasses alternative non-hybrid baseline techniques when it comes to detecting fabricated news. This underscores the efficiency and supremacy of the hybrid model in precisely categorizing instances of fake news. The proposed model showed excellent performance on the ISOT dataset. As the epochs progressed, both the training and testing accuracies increased, while the corresponding loss values decreased. This behavior suggests that the model effectively learned to classify the articles and improve its accuracy over time.

Nevertheless, when analyzing the FA-KES dataset, it's notable that the progression of training and validation accuracy did not display a consistent and gradual rise. This discrepancy is particularly pronounced in the case of the validation accuracy. This indicates that the model faced challenges in generalizing well to the unseen data in the validation set. It suggests that further improvements or adjustments may be required to enhance the model's performance on this specific dataset [9].

One such approach would be to develop an automated method for spotting false information in the Chrome browser environment that is specially designed to recognize misleading material on Facebook. This strategy could involve using classification methods like Support Vector Machine, K Nearest Neighbours, Decision Tree, Naive Bayes, and Logistic Regression. With a high accuracy rate of 99.4%, the deep learning algorithm Long Short-Term Memory (LSTM) outperformed conventional machine learning (ML) methods in terms of performance. The outcome shows that the deep learning strategy achieved higher classification accuracy by successfully learning intricate patterns and representations in the data. The LSTM's (long short-term memory) great accuracy demonstrates its capacity to identify significant characteristics and provide precise predictions in the context of a given task [10].

The linguistic model pulls grammatical, emotive, syntactic, and readability elements from specific news items in addition to other language properties. The created linguistic model detects and categorizes bogus news with an average accuracy of 86% [11].

Within the examination of the public dataset, a comprehensive assortment of 23 supervised A.I. algorithms were employed. Numerous techniques are covered, such as logistic model tree (LMT), weighted instances handler wrapper (WIHW), ridor, multi-layer perceptron (MLP), ordinal learning model (OLM), instance-based classifier (IBk), randomizable filtered classifier (RFC), decision stump, ZeroR, JRip, stochastic gradient descent (SGD), parameters selected through cross-validation (CVPS), Simple Cart, kernel logistic regression (KLR), attribute-selected classifier (ASC), J48, locally weighted learning (LWL), bagging, decision tree, and sequential minimal optimization (SMO)." As part of the investigation, these algorithms were methodically applied to a structured news dataset. According to the outcomes, the Decision Tree technique yielded the highest mean values for accuracy, precision, and F-measure. With 1,000 values, the ZeroR, CVPS, and WIHW algorithms appear to be the optimal ones regarding recall [12].

It is common to observe improved performance across different classifiers when using Long-Short Term Memory (LSTM), Convolutional Neural Networks (CNN), and Capsule Networks–three Deep Learning (DL) models—to implement relational features like named entities, sentiment analysis, and facts extracted from both structured (like Knowledge Graphs) and unstructured (like text) data [13].

A link between Textual Entailment (TE) and the stance detection issue introduced many systems that depend on deep learning (DL), statistical machine learning (ML), and a hybrid strategy integrating both pedagogues. Through empirical review, it has shown promising performance that exceeds the potential of the most advanced system now in use. Additionally, our method performs better than the cutting-edge F1 score that our SVM-based model was able to acquire pre-trained agreement class [14].

Utilizing the NaiveBayes, Support Vector Machine (SVM), and NLP algorithms, natural language processing (NLP) and machine learning (ML) are used to gather news data and determine if it is authentic or fraudulent. With an accuracy level of 93.6%, the suggested model is operating admirably and successfully assessing the accuracy of the outcomes [15].

Three classifiers are suggested, each of which will use a different pre-trained model to incorporate the input news items. Connect a CNN (convolutional neural network), SLP (single-layer perceptron), and MLP (multi-layer perceptron) after the embedding layer. The embedding layer consists of novel pre-trained models such as GPT2, Funnel Transformer, BERT, and RoBERTa. Through the use of these models, we can take use of their deeply contextualized representations, which will help us provide more meaningful and accurate categorization results. Perform analyses on three well-known synthetic news datasets: COVID-19, ISOT, and LIAR to assess the performance and efficacy of suggested models [16].

We use a deep learning methodology in this study, combining several architectures, including ResNet, CNN, and Bi-LSTM. We also include word embeddings that have already been trained using four different datasets. This all-encompassing approach attempts to improve the model's functionality and generalizability, enabling us to provide jobs with more precise and reliable outcomes. The study's findings revealed that the Bidirectional LSTM design demonstrated superior performance compared to CNN and ResNet across all the tested datasets. This shows that the Bidirectional LSTM model performed tasks more accurately and dependably by better capturing the underlying patterns and relationships in the data [17].

When it comes to predicting false news, Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) models have excelled performance benchmarks. The hybrid architecture of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) proves to be a better solution than other solutions like GRU (gated recurrent units), RNN (recurrent neural networks), or feed-forward networks when evaluating the problem of forecasting fake news and incorporating a wider range of input characteristics. This combined strategy outperforms the others with a precision value of 97.21%, demonstrating its potential for more precise forecasts [18].

The reviewed literature has been summarized in TABLE 1 to address the problem in an effective way.

According to the literature review and from TABLE 1 we reached the conclusion that the existing approaches of fake news detection do not effectively detect fake news. To tackle this accurate detection problem we need a more accurate and efficient deep learning model to correctly classify the fake and true news.

III. MATERIALS AND METHODS

The parts that follow go over the basic ideas of the models that have been suggested.

A. DATASET

The whole ISOT dataset was selected from reliable, actual sources [12]. The authentic pieces were obtained by information extraction from the prestigious news website Reuters.com. The false news articles, on the other hand, were put up using Wikipedia and other dubious sources that were uncovered by Politifact, a well-known American fact-checking group. The dataset includes a wide variety of articles on various subjects, with a strong emphasis on political and international news. It has two CSV files with more than 12,600 items in each category.

At the outset file named True.csv contains authentic items from Reuters.com, whereas the subsequent file named Fake.csv contains false news pieces from several questionable news sources. The title, text, type (fake or real), and publication date of each article are all included in the dataset. This dataset consists of 44,898 data elements in total, 21,417 of which are marked as "real news" (labeled with a 1) and 23,481 of which are marked as "fake news" (labeled with a 0). The primary emphasis was on gathering articles from the years 2016 to 2017, in order to guarantee that the data was pertinent to Kaggle.com. The data was cleaned and processed, however, the original punctuation and errors found in the texts of the bogus news were not removed. This dataset serves as a valuable resource for developing and evaluating models for spotting bogus news, providing researchers with real-world data to address the challenges associated with identifying deceptive information from legitimate news articles. TABLE 2 shows the classification of the ISOT dataset.

B. CONVENTIONAL LSTM

Similar to conventional RNNs, a typical LSTM (Long Short-Term Memory) model processes data that conveys information as it advances. The key differences lie in the functions performed inside the LSTM cells, enabling them to retain or discard data. The cell state and the many gates make up the fundamental components of an LSTM. The cell state functions as a channel for the transmission of crucial data during data processing, serving as the network's memory in essence. The gates create separate neural networks by deciding what information is allowed into the cell state. The gates learn which information to keep and which to discard throughout the training phase. Three gates—an input gate, an output gate, and a forget gate—control the information flow in an LSTM cell. The input gate determines which data from the current state should be added. The forget gate decides what needs to be kept from the previous state, while the next concealed state is selected by the output gate. FIGURE 1 illustrates the arrangement of a traditional LSTM. The following equations (1) to (5) can be used to represent the mathematical connections between the inputs and outputs at both time t and time t - 1:

$$i_t = (Wx_i x_t + Wh_i h_{t-1} + Wc_i C_{t-1} + b_t) \tag{1}$$

$$f_t = \left(W_{xf} X_t + W_{hf} h_{t-1} + W_{cf} C_{t-1} + b_f\right) \tag{2}$$

TABLE 1. Overview of existing research.

Ref	Year	Feature extraction and selection algorithm	Data set	Classification Models	No. of classification classes	Evolution Metrics
[4]	2021	TF, TF-IDF	ISOT and KDnugget (fake or real news)	SVM, DT, KNN, RF and LR and three deep learning models LSTM, GRU, CNN	Two classes (fake or real)	On ISOT(accuracy=99.94%) KD-nugget(accuracy=96%)
[7]	2021	GloVe	Liar, Fake or Real news, Combined Corpus	RoBERTa	Two classes (fake or real)	98% accuracy
[8]	2021	Bag of words (BOW)	ISOT and Liar	Ensemble Model(DT,RF and Extra Tree Classifier)	Two classes (fake or real)	ISOT (Training Accuracy 99.8% Testing Accuracy 44.15%) Liar (100% accuracy)
[9]	2021	Embedding (Word2Vec, GloVe)	ISOT and FA-KES	CNN and RNN	Two classes (fake or real)	(Achieving an accuracy of approx.. 60% on the FA-KES dataset comprising 804 articles, and approximately 100% accuracy on the ISOT dataset containing 45,000 articles)
[10]	2021	—	—-	LSTM (Long Short term memory)	Two classes (fake or real)	Accuracy 99.42%
[11]	2021	Linguistic feature extraction	Random Political News Data and Buzzfeed Political News Data	combined linguistic feature-driven	Two classes (fake or Not Fake)	Accuracy 86%
[12]	2020	TF weighting method and Document-Term Matrix	BuzzFeed Political News Data set, Random Political News Data set, ISOT Fake News Data set	The employed techniques encompass a range of algorithms, such as BayesNet, JRip, OneR, Decision Stump, ZeroR, Stochastic Gradient Descent (SGD), Cross-Validation Parameter Selection (CVPS), Randomizable Filtered Classifier (RFC), Logistic Model Tree (LMT), Locally Weighted Learning (LWL), Classification Via Clustering (CvC), Weighted Instances Handler Wrapper (WIHW), Ridor, Multi-Layer Perceptron (MLP), Ordinal Learning Model (OLM), Simple Cart, Attribute Selected Classifier (ASC), J48, Sequential Minimal Optimization (SMO), Bagging, Decision Tree (DT), Instance-Based Classifier (IBk), and Kernel Logistic Regression (KLR).	Two classes (fake or real)	BuzzFeed Political News (J48 highest accuracy), Random Political News(SMO highest accuracy), ISOT (DT highest accuracy)
[13]	2019	Embedding(word2vec, GloVe, fastText)	Liar	LSTM, BasicLSTM, BI-LSTM, GRU, CNN, and CapsNetLSTM.	6classes (False, Barely True, Half True, Mostly true, True, Pants on fire)	accuracy increase of up to 5–6%
[14]	2019	Universal sentence encoder (USE)	Fake or Non-Fake	SVM, MLP, Univ_Sen_Enc, Univ Sen Enc Features, Official Baseline, HUMAN UPPER BOUND	Four classes (Agree, Disagree, Discuss, Unrelated) Univ Sen Enc Features	(Accuracy 82.54%)
[15]	2019	Counter Vectorizer Generator	Fake or Real	SVM, NB, NLP	Two classes(Real or fake)	Accuracy 93.6%
[16]	2021	SLP, MLP,CNN	LIAR, ISOT, and COVID-19	BERT, RoBERTa, GPT2, and Funnel Transformer	Two classes(Real or fake)	Funnel-CNN(Accuracy 99.96% on ISOT) RoBERTa-CNN(Accuracy 97.43% Covid-19) RoBERTa-GN-CNN (Accuracy 97.80%)
[17]	2022	Word2Vec ,GloVe, and fast-Text	ISOT	CNN, BiLSTM, ResNet	Two classes(Real or fake)	BiLSTM+GloVe (Accuracy 99.95%) CNN+fastText(Accuracy 99.89%) ResNet+GloVe (Accuracy 99.91%)
[18]	2020	GloVe	Kaggle	GRU, RNN	Two classes(Real or fake)	(Accuracy 97.21%)
[19]	2024	-	Twitter15, Twitter16, Weibo	SMNet	2 classes	95.9% accuracy on Weibo, 91.7% and 93.5% on Twitter15, Twitter16 respectively

TABLE 2. ISOT dataset overview.

News	No. of Articles	Subjects
Real-News	21417	Type World news: 10145 Political news: 11272
Fake-News	23481	Type Govt. news: 1570 Middle East news: 778 US-news: 783 Left-news: 4459 Politics: 6841 News: 9050

$$c_t = f_t C_{t-1} + i_t \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right) \quad (3)$$

$$o_t = \left(W_{x0}x_t + W_{h0}h_{t-1} + W_{c0}c_{t-1} + b_0\right) \quad (4)$$

$$h_t = o_t tanh\left(c_t\right) \quad (5)$$

In this instance, c stands for the cell's state, σ for the sigmoid function, and tanh for the activation function. The input vector is represented by x, while the outcome is denoted by h_t. W and b stand for the weights and biases parameters, respectively. While i_t (the input gate) supplies new data to the cell's state, the forget gate f_t filters out unwanted information. The output gate, o_t, generates pertinent information.

C. BIDIRECTIONAL LSTM

Recurrent neural network (RNN) models such as Bidirectional Long-Short Term Memory (BiLSTM) improve on standard LSTM models by processing input data in both forward and backward directions [20]. Since of this special feature, the model is very successful for a variety of sequence processing tasks since it can simultaneously gather information from the past and the future environment. In a standard LSTM, information flows through the network in a unidirectional manner, either from past to present (forward LSTM) or from present to future (backward LSTM). However, BiLSTM divides the network into two components: one processes the input sequence from start to end (forward LSTM), and the other processes the sequence from end to start (backward LSTM) [21]. The hidden states from both directions are typically concatenated or summed at each time step, providing a comprehensive representation of the input

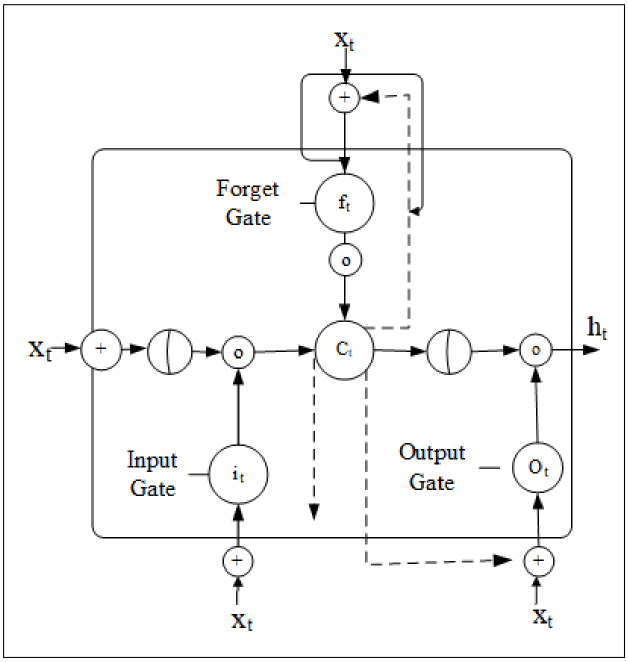


FIGURE 1. Fundamental design of a long short-term memory cell.

sequence. The main benefit of employing a BiLSTM is that it can better comprehend the input sequence by capturing context and long-term relationships from both directions. This feature improves the performance of several natural language processing tasks, such as sentiment analysis, named entity recognition, machine translation, and part-of-speech tagging. FIGURE 2 illustrates the design of BiLSTM.

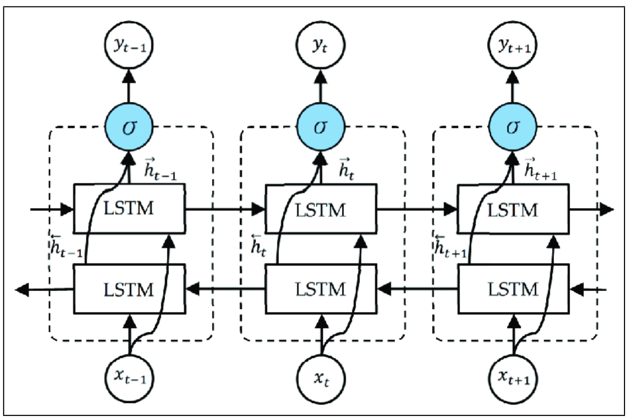


FIGURE 2. Architecture of Bidirectional LSTM.

D. SELF ATTENTION

Within the realm of computer vision, this approach has demonstrated notable effectiveness in tasks such as image classification, object recognition, and segmentation. The self-attention process, allows a prototype to selectively concentrate on essential components of its input data. In our study, the neural network architecture's self-attention mechanism was incorporated as a distinct layer. The self-attention process begins by transforming the input feature map into 3 distinct matrices 1 (a query matrix), 2 (a key matrix), and 3 (a value matrix), when provided with an input feature map having dimensions B C (where B denotes batch size and C represents the number of channels). With a kernel size of 1, convolutional layers are used to create these matrices and subsequently reshaped into matrices with a size of B (C/8). The energy between the query and key matrices is calculated using the dot product within the attention mechanism. The softmax function is used to the resultant energy matrix to provide attention weights that capture the importance of each feature for the job at hand. The value matrix is then weighted using these attention weights, producing the ultimate feature map weighted by attention. The attention-weighted feature map is integrated with the real input feature map to complete the description, and the resulting feature map is transmitted through the remaining network layers.

The network can dynamically alter its processing focus due to the integration of the self-attention mechanism, which enhances performance for a variety of computer vision applications. The self-attention process is mathematically depicted by the following equations (6) and (7):

$$attention = Softmax(query * key) \tag{6}$$

$$out = attention * value \tag{7}$$

where the matrices represent the query, key, and value. attention and out are the corresponding attention weights and final output of the self-attention mechanism [5], [22], [23].

E. BIDIRECTIONAL LSTM (BILSTM)+SELF ATTENTION

The performance of conventional LSTMs in classification tasks is improved by the introduction of bidirectional LSTMs (BiLSTM). BiLSTM involves simultaneously training 2 LSTMs on the provided data. While the other LSTM processes the provided data that has been inverted, one LSTM processes the original input data. This strategy gives the network access to more contextual data and may produce quicker, more precise answers. The concept behind BiLSTM is really simple. Replicating the initial recurrent layer in the network involves providing the original input data to the first layer, and simultaneously feeding the duplicated layer with a reversed replica of the input. This method successfully tackles the vanishing gradients problem that happens in conventional RNNs. A bidirectional LSTM (BiLSTM) model utilizes all of the previous and upcoming input data during the course of the training process. From L2R (employing a forward-facing hidden layer) and R2L (employing a backward-facing hidden layer), respectively, is how BiLSTM inherently analyzes the input data [24].

To implement BiLSTMs in Python, a bi-directional layer wrapper is provided by the Keras package. This wrapper makes it easier to create a bidirectional architecture by accepting the first LSTM layer as an input. One of the crucial factors that may be controlled is the merge mode, which controls how the outputs from the forward and backward LSTMs are merged and mixed before being relayed to the next layer in the network. The merging mode can be chosen based on the particular needs of the work at hand. Using the Keras library, a BiLSTM's information structure and flow are shown in FIGURE 3.

The output (y), forward hidden layer and backward hidden layer are calculated by Equations (8), (9), and (10) respectively [24], [25].

$$h_t^{\vec{t}} = H\left(W_x\vec{h}x_t + W_h\vec{h}h_{t-1}h^{\rightarrow}\right) \tag{8}$$

$$h_t^{\leftarrow} = H\left(W_{x\leftarrow}x_t + W_h^{\leftarrow}ht + \overset{\leftarrow}{\leftarrow} h^{\leftarrow} + b_h^{\leftarrow}\right) \tag{9}$$

$$y_t = W_h \rightarrow {}_y h_t + W_{hy}^{\leftarrow} t + b_y \tag{10}$$

where W stands for weight matrices (Wxh → and Wxh ←, respectively, represent the weight matrices for the forward input-hidden and backward input-hidden connections, b (bh →and bh ←) stands for bias vectors in both directions and the term represents the hidden layer. Additionally, the BiLSTM was enhanced through the introduction of a self-attention process. An embedded layer in the BiLSTM model learns the dense vector representations (embeddings) for every word within the input sequence from a fixed-length series of words written as integers (one-hot encoded). The word's semantic meaning is captured by the embeddings. Prior to being stored, the result of the LSTM layer is processed by an attention layer. The attention weight of each input sequence element is determined by the Attention layer. The weight of each word indicates its importance in the context of the entire sequence. The attention weights, which are determined using a trainable weight matrix, are then applied to the LSTM outputs.

To avoid overfitting, a Dropout layer is added after the Attention layer. Model regularization is facilitated by dropout, wherein a fraction of input units are randomly set to 0 during training. The Dropout layer's output is coupled to a dense layer with sigmoid activation. This layer's prediction determines whether the input sequence falls within the category of false news or not (0 or 1). Various assessment measures are used to assess the suggested model. Furthermore, held-out validation will be employed for both training and testing the model. In FIGURE 3 with rectangular boxes denoting the layers, arrows denoting data flow, and the model's pseudocode in algorithm 1 to illustrate the suggested Self Attention+BiLSTM model.

F. METRICS OF PERFORMANCE EVALUATION OF THE PROPOSED MODEL

Popular evaluation metrics such as Accuracy (Acc), Specificity, sensitivity (Sn), Recall, Precision, F1-Measure, and AUC-ROC are incorporated for the evaluation of the model. For binary classification problems, the confusion matrix is given in TABLE 3.

According to TABLE 3, TP, TN, FP, and FN are defined as follows:

TABLE 3. Confusion matrix for binary classification problem.

	anticipating Fake news subject	anticipating True news subject
Real fake news topic	TP	FN
Real true news topic	FP	TN

- TP (True Positive) if the topic is accurately identified as fake news.
- If a topic of true news is accurately categorized as true news, the result is TN (True Negative).
- FP (False positive) occurs when a topic of True news is mistakenly labeled as Fake news.
- If false information is mistakenly categorized as true, it is a False Negative (FN). These measures for performance evaluation are theoretically stated and represented quantitatively in equations 11-15 respectively.
- Accuracy (Acc): Accuracy is the metric used to assess the overall effectiveness of the categorization system.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{11}$$

- Sensitivity and Recall (Sn/Rc): It proved that the news is false and the forecast is correct.

$$Sn/Rc = \frac{TP}{TP + FN} \times 100\% \tag{12}$$

- Specificity (Sp): It shows that the issue is true news and the prognosis is negative.

$$Sp = \frac{TN}{TP + FP} \times 100\% \tag{13}$$

- Precision (Pr): The model's precision is the proportion of accurate outcomes to all results produced by the model.

$$Pr = \frac{TP}{TP + FP} \times 100\% \tag{14}$$

- F1-score: The F1-score represents the harmonic mean of recall and accuracy.

$$F1 - Score = 2 \times \frac{Precision \times recall}{Precision + recall} \tag{15}$$

- Area under the receiver operating characteristics curve (AUC-ROC): The AUC-ROC is a widely used statistic in model evaluation.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. EXPERIMENTALS SETUP

Several experiments will be conducted to evaluate the performance of the proposed model (SA BiLSTM). The ISOT dataset will initially be split into training and testing segments in order to train and evaluate the suggested model. We used a hold-back approach, splitting the dataset across all trials into 30% for testing and 70% for training, to ensure dependable findings. For training parameter optimization, The studies were conducted across 140 epochs using the ADAM and SGD algorithms, with a batch size of 120 and a mini-batch size

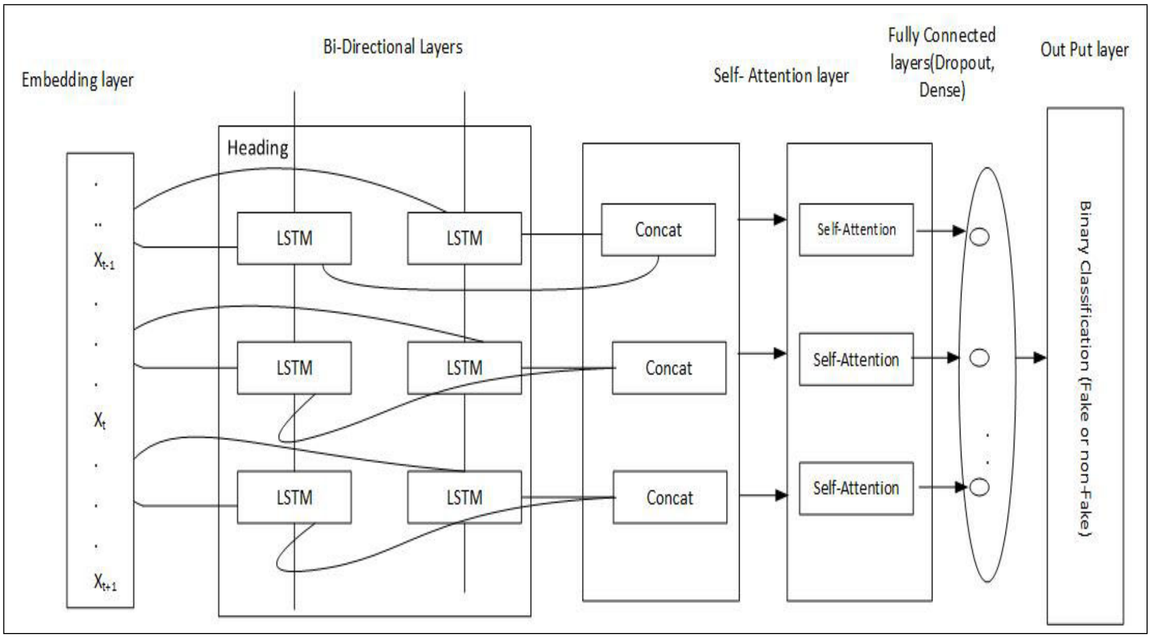


FIGURE 3. Self Attention +BiLSTM (SA-BiLSTM) model for Binary classification (Fake or Non-Fake News).

of 9. We also set the learning rate (LR) α to 0.0001. In all the tests, sigmoid is employed as the outer activation function of the model, while Relu serves as the inner activation function. For each experiment conducted, the hardware setup consisted of a computer operating on Windows 8, outfitted with a CPU and GPU. Python v3.7 is a requirement for all experiment software. As a high-level API, Keras v2.2.4 will be used to implement the LSTM, BiLSTM, and Self-Attention+BiLSTM models, using TensorFlow v1.12 as the back-end software. To ensure reliable and consistent outcomes, we have iterated three to five times in each experiment to obtain stable results.

B. RESULTS AND ANALYSIS

1) DATA GATHERING AND PREPOSSESSION PROCEDURES

44,898 data elements make up the ISOT dataset; 21,417 of these are classified as genuine news (labeled as 1), and 23,481 as false news (labeled as 0).

2) CNN ARCHITECTURE OUTCOMES ON THE ISOT DATASET

CNN's performance was assessed on the ISOT dataset. The hyperparameter setup includes the ADAM and SGD optimizers (OP) with a learning rate (LR) of 0.0001. The suggested Bi-LSTM+Attention yielded results that we compared using two different optimization techniques. The training period was set at 140 and the batch size at 120. A range of assessment indicators was employed to evaluate the performance of the model. To evaluate the model's performance, a number of evaluation measures were applied.

The experimental findings and all of the hyperparameters are listed in TABLE 4. TABLE 4 demonstrates that the CNN architecture with the SGD optimizer and at a learning rate of 0.0001 obtained 90.22% accuracy, 92.77% specificity, 92.88% sensitivity, 99.23% precision, 98.22% F1 score, and 95.95% AUC. Utilizing the ADAM optimizer and operating at a learning rate of 0.0001 CNN obtained 92.47% accuracy, 95.80% specificity, 95.56% sensitivity, 99.40% precision, 98.09% F1 score, and 96.93% AUC. Results of CNN architecture on the ISOT dataset are shown graphically in Figure 4.

TABLE 4. Results of CNN on ISOT dataset.

Parameters		Metrics					
Optimizer	LR	Ac(%)	Sp(%)	Sn(%)	Pr(%)	F1(%)	AUC(%)
SGD	0.0001	90.22	92.77	92.88	99.23	98.22	95.95
ADAM	-	92.47	95.80	95.56	99.40	98.09	96.93

3) GRU ARCHITECTURE OUTCOMES ON THE ISOT DATASET

Two optimization algorithms were strategically leveraged to juxtapose the outcomes of the innovatively proposed Bi-LSTM+Attention architecture. The prescribed experimental configuration specified a batch size of 120 and a training epoch of 140. Diverse and sophisticated evaluation metrics were meticulously deployed to scrutinize and gauge the model's intricate performance landscape. The evaluative framework extended to the assessment of GRU [26] performance, utilizing the ISOT dataset. The hyperparameter configuration encompassed the judicious selection of ADAM and SGD optimizers (OP) concomitant with a nuanced learning rate (LR) set at 0.0001.

The experimental findings and all of the hyperparameters are listed in TABLE 5. TABLE 5 shows that the GRU

Algorithm 1 Self-Attention + BiLSTM (SA-BiLSTM) Model

Input: Dataset D, Feature X, Target label Y, Learning rate η, Number of epochs E, Bidirectional LSTM layer L, Self-attention layer S
Output: Binary Classifications (Fake=1 or Non-Fake=0)
Step 1: Load Dataset (Fake News)
Step 2: Pre-process dataset (Define vocabulary size and sequence length)
    voc_size = 10000
    sent_length = 1000
    One-hot encode the sentences
    Features extraction (x)
    onehot_repr = [one_hot(words, voc_size) for words in corpus]
    Pad sequences
    embedded_docs = pad_sequences(onehot_repr, padding='pre', maxlen=sent_length)
Step 3: Split the data into train and test sets
Step 4: Define the Attention layer
Step 5: Build the model with Attention
    Embedding_vector_features = 100
    model = Sequential()
    model.add(Embedding(voc_size, embedding_vector_features, input_length=sent_length))
    model.add(Bidirectional(LSTM(100, return_sequences=True)))
    model.add(Attention())
    model.add(Dropout(0.3))
    model.add(Dense(1, activation='sigmoid'))
Step 6: Compile the model
Step 7: Train the model
Step 8: Validate model with test data
Step 9: Evaluate model evaluation metrics

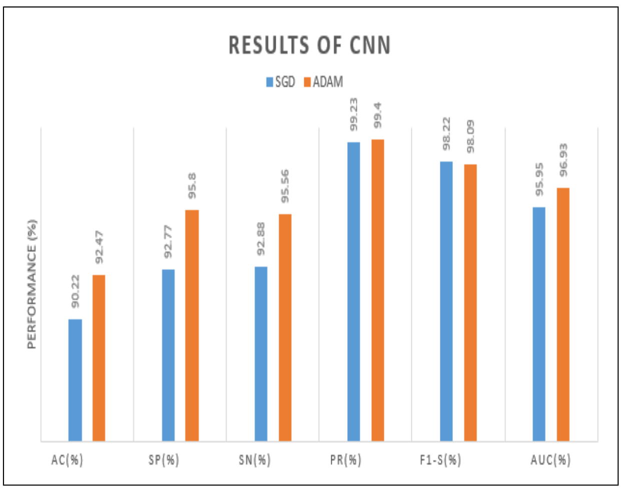


FIGURE 4. Performance of CNN on ISOT dataset.

architecture with the SGD optimizer and at a learning rate of 0.0001 obtained an accuracy of 95.30%, specificity of 97.44%, sensitivity 96.60%, precision 98.56%, F1 score of 98.50% and AUC 99.07%. The GRU architecture is instantiated with the ADAM optimization algorithm, operating at a learning rate (LR) of 0.0001 obtained 95.52% accuracy, 99.40% specificity, 100.00% sensitivity, 98.98% precision, 98.89% F1 score, and 99.49% AUC. Results of GRU architecture on the original ISOT dataset are shown graphically in FIGURE 5.

TABLE 5. GRU performance on ISOT dataset.

Parameters		Metrics					
Optimizer	LR	Ac(%)	Sp(%)	Sn(%)	Pr(%)	F1(%)	AUC(%)
SGD	0.0001	95.30	97.44	96.60	98.56	98.50	99.07
ADAM	-	95.52	99.40	100.00	98.98	98.89	99.49

4) OUTCOMES OF THE ATTENTION-GRU ARCHITECTURE, ON THE ISOT DATASET

Using the ISOT dataset, Attention-GRU's performance was assessed. The ADAM and SGD optimizers (OP) were implemented, both synchronized with a shared learning rate (LR) of 0.0001 are part of the hyperparameter configuration. The suggested Bi-LSTM+Attention yielded results that we compared using two different optimization techniques. The training period was set at 140 and the batch size at 120. A range of assessment indicators were employed in order to assess the performance of the model. To evaluate the model's performance, a number of evaluation measures were applied.

The experimental findings, along with all the hyperparameters, are detailed in TABLE 6. Analysis of TABLE 6 reveals that the Attention-GRU architecture, employing the SGD optimizer with a learning rate of 0.0001, achieved

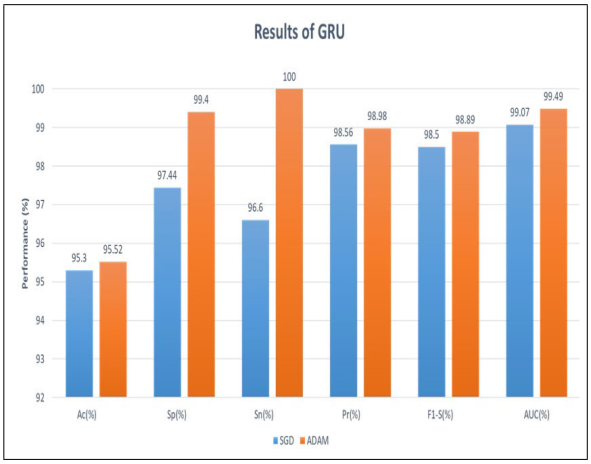


FIGURE 5. Performance of GRU on ISOT dataset.

97.67% accuracy, 100.00% specificity, 98.91% sensitivity, 96.11% precision, 99.00% F1 score, and 98.02% AUC. Meanwhile, the Attention-GRU architecture, utilizing the ADAM optimizer with a LR of 0.0001, demonstrated superior performance with 98.46% accuracy, 99.78% specificity, 99.00% sensitivity, 100.00% precision, 99.78% F1 score, and 99.59% AUC. Graphical representation of the Attention-GRU architecture results on the original ISOT dataset is illustrated in FIGURE 6.

TABLE 6. attention GRU performance on ISOT dataset.

Parameters		Metrics					
Optimizer	LR	Ac(%)	Sp(%)	Sn(%)	Pr(%)	F1(%)	AUC(%)
SGD	0.0001	97.67	100.00	98.91	96.11	99.00	98.02
ADAM	-	98.46	99.78	99.00	100.00	99.78	99.59

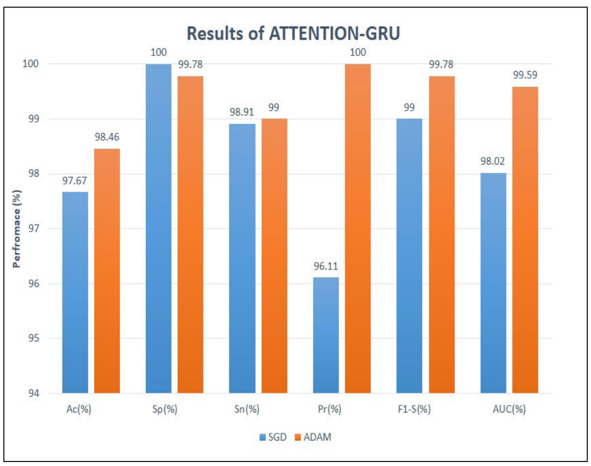


FIGURE 6. Performance of Attention-GRU on ISOT dataset.

5) OUTCOMES OF THE ISOT DATASET USING THE LSTM ARCHITECTURE

Using the ISOT dataset, LSTM performance was assessed. The ADAM and SGD optimizers (OP) were employed, both with a shared learning rate (LR) of 0.0001 are part of the hyperparameter configuration. The suggested Bi-LSTM+Attention yielded results that we compared using two different optimization techniques. The training period was set at 140 and the batch size at 120. A range of assessment indicators was employed to evaluate the performance of the model. To evaluate the model's performance, a number of evaluation measures were applied.

The experimental findings and comprehensive hyperparameter details are outlined in TABLE 7. Examination of TABLE 7 reveals that the LSTM architecture, utilizing the SGD optimizer with a learning rate of 0.0001, achieved 98.89% accuracy, 97.03% specificity, 98.45% sensitivity, 97.04% precision, 98.99% F1 score, and 98.45% AUC. In comparison, the LSTM architecture, employing the ADAM optimizer with a learning rate of 0.0001, demonstrated notable performance with 98.94% accuracy, 99.02% specificity, 96.34% sensitivity, 99.23% precision, 99.10% F1 score, and 99.02% AUC. Graphical representation of the LSTM architecture results on the ISOT dataset is depicted in FIGURE 6.

TABLE 7. LSTM performance on ISOT dataset.

Parameters		Metrics					
Optimizer	LR	Ac(%)	Sp(%)	Sn(%)	Pr(%)	F1(%)	AUC(%)
SGD	0.0001	98.89	97.03	98.45	97.04	98.99	98.45
ADAM	-	98.94	99.02	96.34	99.23	99.10	99.02

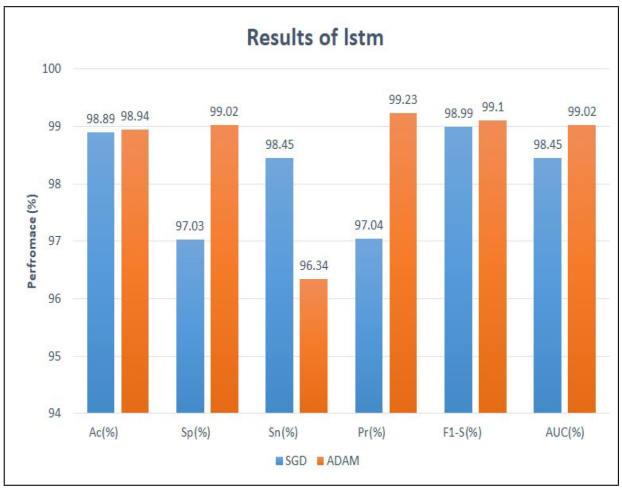


FIGURE 7. Performance of LSTM on ISOT dataset.

6) RESULTS OF THE BI-LSTM ARCHITECTURE, ON THE ISOT DATASET

Using the ISOT dataset, Bi-LSTM performance was assessed. The inclusion of the ADAM and SGD optimizers (OP) in the hyperparameter configuration is marked by a shared learning rate (LR) of 0.0001. The suggested Bi-LSTM+Attention yielded results that we compared using two different optimization techniques. The training period was set at 140 and the batch size at 120. A range of assessment indicators were employed In order to assess the model's performance. To evaluate the model's performance, a number of evaluation measures were applied.

The experimental results, along with a comprehensive list of hyperparameters, can be found in TABLE 8. The TABLE 8 illustrates that the Bi-LSTM architecture, when utilizing the SGD optimization technique with a learning rate (LR) of 0.0001, achieved a performance of 99.12% accuracy, 99.10% specificity, 97.35% sensitivity, 99.46% precision, 99.45% F1 score, and 98.89% AUC. Furthermore, employing the ADAM optimization method with an LR of 0.001 resulted in an even higher performance for the Bi-LSTM architecture, attaining 99.34% accuracy, 100% specificity, 98.68% sensitivity, 99.67% precision, 99.78% F1 score, and 99.01% AUC. Graphical representation of the Bi-LSTM architecture results on the ISOT dataset is presented in FIGURE 7.

TABLE 8. Bi LSTM performance on ISOT dataset.

Parameters		Metrics					
Optimizer	LR	Ac(%)	Sp(%)	Sn(%)	Pr(%)	F1(%)	AUC(%)
SGD	0.0001	99.12	99.10	97.35	99.46	99.45	99.89
ADAM	-	99.34	100.00	98.68	99.67	99.78	99.01

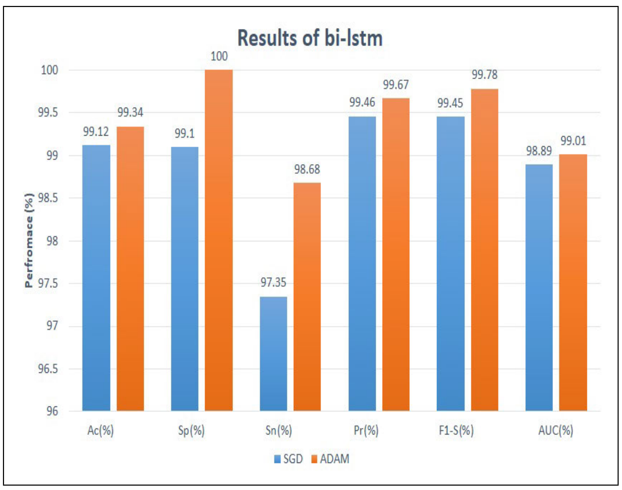


FIGURE 8. Performance of Bi-LSTM on ISOT dataset.

7) OUTCOMES OF THE PRESENTED BI-LSTM ARCHITECTURE WITH ATTENTION

Using the ISOT dataset, Bi-LSTM with Attention performance was assessed. The ADAM and SGD optimizers (OP), both configured with a learning rate (LR) of 0.0001 are part of the hyperparameter configuration. The suggested Bi-LSTM+Attention yielded results that we compared using two different optimization techniques. The training period was set at 140 and the batch size at 120. A range of assessment indicators were employed in order to assess the performance of the model. To evaluate the model's performance, a number of evaluation measures were applied.

The experimental results, along with a detailed list of hyperparameters, can be found in TABLE 9. The TABLE 9 highlights that the Bi-LSTM+Attention architecture, when utilizing the SGD optimization technique with a learning rate (LR) of 0.0001, achieved a remarkable performance of 99.67% accuracy, 100% specificity, 99.02% sensitivity, 99.89% precision, 99.79% F1 score, and 99.80% AUC. Furthermore, employing the ADAM optimization method with an LR of 0.0001 resulted in even higher performance for the Bi-LSTM+Attention architecture, attaining 99.98% accuracy, 99.97% specificity, 99.97% sensitivity, 99.56% precision, 99.96% F1 score, and 99.98% AUC. Graphical representation of the Bi-LSTM+Attention architecture results on the ISOT dataset is presented in FIGURE 8.

It is noteworthy that our proposed model exhibited superior results when trained on the ISOT dataset, particularly with the ADAM optimizer and a learning rate of 0.0001. The obtained findings were 99.98% accuracy, 99.97% specificity, 99.97% sensitivity, 99.56% precision, 99.96% F1 score, and 99.98% AUC, showcasing enhanced performance across all evaluation metrics.

TABLE 9. Proposed SA+Bi-LSTM performance on ISOT dataset.

Parameters		Metrics					
Optimizer	LR	Ac(%)	Sp(%)	Sn(%)	Pr(%)	F1(%)	AUC(%)
SGD	0.0001	99.67	100.00	99.02	99.89	99.79	99.80
ADAM	-	99.98	99.97	99.97	99.56	99.96	99.98

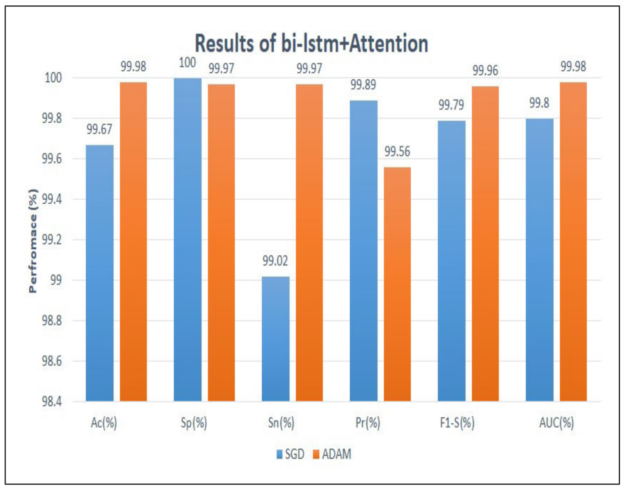


FIGURE 9. Performance of the Proposed model SA-BiLSTM on ISOT dataset.

8) MODELS (CNN, GRU, SA-GRU, LSTM, BI-LSTM, SA-BI-LSTM) PERFORMANCE COMPARISON ON ISOT DATASET

The performance juxtaposition of various models vis-Ã -vis the recommended SA+Bi-LSTM model is

TABLE 10. Models (CNN, GRU, SA-GRU, LSTM, Bi-LSTM, SA-Bi-LSTM) performance comparison on ISOT dataset. The space complexity of each model is the number of trainable parameters. M = Million. The space complexity increases with the increasing number of trainable parameters. The time complexity is the training time (in hours) of the models.

Model	Parameters	Space Complexity	Time Complexity	Metrics					
	Optimizer			Ac(%)	Sp(%)	Sn(%)	Pr(%)	F1(%)	AUC(%)
CNN	SGD	26.4M	3.44h	90.22	92.77	92.88	99.23	98.22	95.95
	ADAM			92.47	95.80	95.56	99.40	98.09	96.93
GRU	-	120.5M	6.25h	95.30	97.44	96.60	98.56	98.50	99.07
	-			95.52	99.40	100.00	98.98	98.89	99.49
SA-GRU	-	134.04	7.13h	97.67	100.00	98.91	96.11	99.00	98.02
	-			98.46	99.78	99.00	100.00	99.78	99.59
LSTM	-	37M	4.03h	98.89	97.03	98.45	97.04	98.99	98.45
	-			98.94	99.02	96.34	99.23	99.10	99.02
Bi-LSTM	-	41M	4.35h	99.12	99.10	97.35	99.46	99.45	99.89
	-			99.34	100.00	98.68	99.67	99.78	99.01
SA-Bi-LSTM	-	51.09M	7.25h	99.67	100.00	99.02	99.89	99.79	99.80
	-			99.98	99.97	99.97	99.56	99.96	99.98

meticulously delineated in TABLE 10. In this comprehensive comparative scrutiny, the recommended model manifests its preeminence by attaining superlative outcomes across an array of metrics, including F1-score, AUC, accuracy, specificity, sensitivity, and precision. Specifically, under the aegis of the SGD optimizer, the recommended model attains a pinnacle of 99.67% accuracy, 100% specificity, 99.02% sensitivity, 99.89% precision, 99.79% F1-score, and 99.80% AUC. Furthermore, leveraging the ADAM optimizer, the model SA-Bi-LSTM obtained 99.96% F1-score, 99.98% AUC, 99.98% accuracy, 99.97% specificity, 99.97% sensitivity, and 99.56% precision. These findings underscore the unparalleled efficacy of the recommended SA+Bi-LSTM model, positioning it as a paragon of performance in relation to its counterparts in the evaluative spectrum.

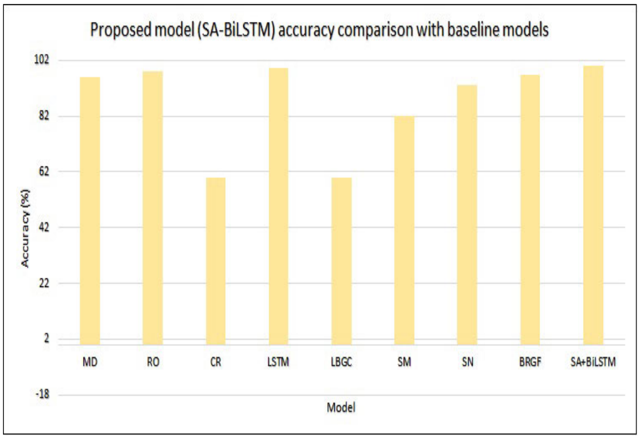


FIGURE 10. Proposed model (SA-BiLSTM) accuracy comparison with baseline models with LR=0.0001.

9) PERFORMANCE COMPARISON OF THE PROPOSED METHOD WITH THE BASELINE METHODS IN TERMS OF ACCURACY

In TABLE 11, a comparative analysis of the baseline model's accuracy is conducted in juxtaposition with the proposed SA-BiLSTM model. Notably, the proposed model attains a superior level of accuracy in contrast to established models. The proposed model achieves state-of-the-art performance, particularly in the realm of accuracy. The graphical representation of this accuracy comparison is vividly illustrated in FIGURE 10.

10) SPACE AND TIME COMPLEXITY

The space and time complexity are reported in Table 10of the different proposed models (CNN, GRU, SA-GRU, LSTM, Bi-LSTM, SA-Bi-LSTM) used in the prediction of fake news detection. Also in Table 11 we compared the space and time complexity of the proposed model with baseline models. Since the proposed models are convolutional deep learning models, the space complexities are computed in terms of each model's trainable parameters. For the time complexity, the model's training time is used. It could be deduced from Table 10 that SA-GRU has the worst space complexity since its trainable parameter is 134.04 million, while LSTM has the best space-time complexity. Additionally, for the time complexity, the CNN model has the worst time complexity because its training time is 3.44h.

We were unable to determine the models' complexity in terms of algorithmic run-time through experimentation since it was impossible to obtain the models of the rival models in Table 11. Due to the large number of parameters and matrix processing that come with the model's architecture, it is more likely that practically all models using deep learning techniques—convolutional neural networks—will have a worse space and time complexity. Our suggested model outperforms all other approaches in terms of accuracy, even when considering the worst-case time and space complexity. The time complexity is the training time (in hours) of the models as reported in Table 10. The space and time complexity of our model are $\mathcal{O}(cwh+1)f$ and $\mathcal{O}(f*u*m)$ respectively.

11) DISCUSSION

From the experimental results, we have some interesting findings that the SA-Bi-LSTM model's interpretability in

TABLE 11. Comparative Analysis: Evaluating the Proposed Approach Against Baseline Methods. c = the number of convolutional channels, h =height of input, w = width of input, f = the convolutional kernel size, n = the number data samples, k = the number of output neurons, m = the number of input neurons and d = the dimension or feature of the input, K =number of nearest neighbors, $u = c * w * h$.

Reference	Method	Accuracy (%)	Space Complexity	Time Complexity
[10]	LSTM	99.42	$\mathcal{O}(cwh+1)f$	$\mathcal{O}(f*u*m)$
[27]	NSEP	86.8	$\mathcal{O}(cwh+1)f$	$\mathcal{O}(f*u*m)$
[28]	RF	97	$\mathcal{O}(cwh+1)f$	$\mathcal{O}(f*u*m)$
[29]	ML	82.85	$\mathcal{O}(cwh+1)f$	$\mathcal{O}(f*u*m)$
[30]	DL	98.41	$\mathcal{O}(cwh+1)f$	$\mathcal{O}(f*u*m)$
[13]	LBGC	60	$\mathcal{O}(cwh+1)f$	$\mathcal{O}(f*u*m)$
[14]	SM	82	$\mathcal{O}((nd+d^2)+(cwh+1)$	K(nd+Kn)
[31]	Bi-LSTM	92.7	$\mathcal{O}(cwh+1)f$	$\mathcal{O}(f*u*m)$
[15]	SVM&NB	93	$\mathcal{O}((nm+mk)+(n*d))$	$\mathcal{O}((cwh)+(nd+Kn))$
[16]	BRGF	97	$\mathcal{O}((nd+d^2)$	$\mathcal{O}((cwh)+(nd+Kn))$
Proposed model 2024	**SA-BiLSTM**	**99.98**	$\mathcal{O}(cwh+1)f$	$\mathcal{O}(f*u*m)$

```
Attention Weights Shape: (6864, 200)
Example Attention Weights for the first sequence: [ 0.02159311  0.05429225 -0.17641464  0.08242103  0.6672851   0.27314186
 -0.2875364  -0.4614616   0.02080903  0.58964103 -0.37608385  0.24462697
  0.07932809 -0.30291304  0.25243032  0.12944256  0.13159257  0.22275013
 -0.21370572  0.298361   -0.6228883  -0.03117373 -0.03232942 -0.5029683
  0.541561    0.00849815  0.476856   -0.08069775  0.3092471   0.30456194
  0.40587863 -0.12257856  0.7169827   0.23543796  0.02628413 -0.11844114
  0.27301782  0.02257097  0.10838929  0.00774711  0.22794873 -0.41970098
  0.32470372 -0.37921038  0.09663403  0.02319263 -0.52898306 -0.02578744
  0.09008627 -0.37600496  0.10113093 -0.55490583  0.28590906  0.27126947
 -0.08140519 -0.46209186  0.11883277  0.5986585  -0.16311178 -0.48129466
 -0.22984622  0.14259101 -0.29896218 -0.5514946  -0.26607186  0.01317342
  0.48095325  0.33329785  0.08976181  0.1735468  -0.22439551 -0.2347373
 -0.47786826 -0.09157485 -0.12828821 -0.07276954  0.38072032  0.19881451
 -0.26916692 -0.02288418 -0.36132896 -0.07870197 -0.38266814  0.3686719
 -0.4010203  -0.35622832 -0.4804213   0.6766323  -0.18470654  0.48883608
  0.36992073  0.66162026 -0.07521779  0.17597613 -0.00845846 -0.17447428
  0.03124011  0.03122126 -0.3152696   0.20311955  0.0472325   0.41548166
  0.45824075  0.46922594 -0.46109316  0.37357652  0.36153433 -0.15023147
  0.2882235   0.11943583 -0.17444131 -0.15830924  0.20872045  0.39729363
 -0.4394954  -0.20228978  0.26346895 -0.00612427 -0.45983225 -0.22000349
  0.06544428  0.43519402  0.14228362 -0.03710251 -0.00375936 -0.36048698
  0.20401555 -0.16539878  0.49320367  0.01647356 -0.48482633  0.2112809
 -0.0567595  -0.30384195 -0.10197426 -0.06434035  0.32095495 -0.3863763
  0.23046508 -0.32860395  0.00845553 -0.01249738 -0.03814964  0.08803622
 -0.3604308  -0.01584965  0.34662798 -0.44905278  0.4487951   0.07823145
  0.04786896  0.3710562  -0.11606546 -0.0628669   0.3802821  -0.4181252
 -0.00567658 -0.49016052 -0.10805509 -0.11665281  0.44025663  0.32125065
 -0.27661055 -0.31193516 -0.3154511  -0.21432075 -0.43954968 -0.20532675
  0.22641066 -0.28301078 -0.0670559  -0.09774996 -0.27858353 -0.04343012
  0.15384689 -0.5839414   0.39774135 -0.4452243  -0.24418123  0.31307322
 -0.33206862  0.19716066  0.42158902 -0.2532979   0.13364522 -0.4638918
 -0.3427751  -0.01038612 -0.4118546  -0.25174338 -0.33126912  0.37520093
 -0.5494586   0.11092708 -0.107173   -0.23829368 -0.05306242  0.24207997
  0.07049819 -0.47892526]
```

FIGURE 11. Attention weights shape.

fake news detection is driven by its attention mechanism. Attention weights highlight the importance of individual words in the data. The attention mechanism can identify the fake specific features related to the outcomes, which can not only help the domain experts understand the prediction outcomes, but also support the decision makers to make individualized decisions. It can also obtain the important feature in the data by counting the frequency of the top-ranked features in all Fake news. Due to the attention mentioned the proposed model obtained higher predictive performance.

FIGURE 12. The heat map showing the contribution (attention weight) of each feature for readmission identified by the proposed model for randomly selected Fake news.

The experimental findings and all of the hyperparameters are listed in TABLE 10. According to TABLE 10, the Bi-LSTM+Attention architecture achieved 99.67% accuracy, 100% specificity, 99.02% sensitivity, 99.89% precision, 99.79% F1 score, and 99.80% AUC when using the SGD optimization technique with an LR of 0.0001. With the utilization of the ADAM optimization method and a learning rate (LR) of 0.0001, the Bi-LSTM+Attention architecture yielded remarkable results, achieving 99.98% accuracy, 99.97% specificity, 99.97% sensitivity, 99.56% precision, 99.96% F1 score, and 99.98% AUC. The visual representation of these outcomes is presented in FIGURE 3, showcasing the efficacy of the Bi-LSTM architecture with Attention on the ISOT dataset.

In the context of the ISOT dataset, our proposed model surpassed the performance of the ADAM optimizer with a learning rate of 0.0001. The achieved metrics were 99.98% accuracy, 99.97% specificity, 99.97% sensitivity, 99.56% precision, 99.96% F1 score, and 99.98% AUC, underscoring the model's heightened performance. Notably, the empirical evidence suggests that training the model on the ISOT dataset significantly enhances the efficacy of the proposed Bi-LSTM with Attention mechanism, demonstrating improved performance across all evaluation metrics for both optimization methods, SGD and ADAM.

The study leverages a dataset comprising 44,898 data entries from the ISOT dataset, categorized into 21,417 entries of genuine news and 23,481 entries of false news. The performance comparison of various models, including the suggested SA+BiLSTM model, is detailed in TABLE 11. The comparable models were not as well as our model SA-BiLSTM in terms of accuracy.

The experimental evaluation of the BiLSTM with Attention model yielded empirical results that provide deep technical knowledge of its notable effectiveness in the field of fake news identification. The model's architectural intricacies, notably featuring bidirectional long short-term memory (BiLSTM) units and attention mechanisms, contribute synergistically to its superior discriminative capacity. This advanced neural network design facilitates a nuanced encoding of sequential dependencies and contextual nuances, surpassing the limitations of conventional architectures. In the SA-Bi-LSTM model the attention weights shape [6864, 200] for the first sequence is shown in figure 11. The heat map showing the contribution (attention weight) of each feature for readmission identified by the proposed model for randomly selected Fake news as shown in figure 12.

The model's performance transcendence is not confined to surface-level accuracy metrics; rather, it manifests in its adeptness at discerning subtle semantic patterns indicative of misinformation. The model's heightened granularity in capturing intricate features positions it as a robust solution amidst the escalating sophistication of misinformation tactics employed by adversarial entities.

In the digital landscape, the BiLSTM with Attention model emerges as a potent instrument for the restoration of trust in credible information sources. Its discriminative acumen serves as a reliable means to differentiate between authentic and deceptive content, substantiating its efficacy in combating the pervasive spread of misinformation. The meticulous comparative analysis against extant methodologies accentuates the model's unique strengths, positioning it as a frontrunner in the ongoing battle against information distortion.

The BiLSTM with Attention (SA-Bi-LSTM) model breaks from traditional models when sophisticated neural network structures are examined closely. By drawing attention to key terms and contextual details, attention mechanisms play a crucial part in improving interpretability and offer insightful information to decision-makers and domain experts who are trying to understand the model's reasoning behind the decisions made. Furthermore, the space and time complexity of the proposed model SA-Bi-LSTM are $\mathcal{O}(cwh + 1)f$ and $\mathcal{O}(f * u * m)$ respectively.

V. CONCLUSION

Our study showcases the effectiveness of the SA-BiLSTM model in identifying fake news, leveraging attention mechanisms within a Bi-LSTM neural network. The integration of attention mechanisms significantly improved the model's focus on crucial text segments, facilitating the extraction of essential features for precise classification.

Ablation studies underscored the pivotal role of the attention mechanism in identifying textual patterns associated with fake news, providing valuable insights into data processing.

The SA-BiLSTM model demonstrated competitive performance, achieving 99.98% accuracy, surpassing contemporary fake news detection techniques. Its robustness in generalizing to unseen data, validated through cross-validation, highlights its adaptability to diverse news articles. The use of a meticulously annotated dataset by human fact-checkers further enhances the model's learning capabilities.

In the broader context, our research contributes to the ongoing battle against misinformation, aiming to preserve information integrity and enhance public trust in credible sources. Recognizing the continuous and evolving challenge of misinformation, our work emphasizes the need for ongoing research and advancements.

Our research provides a robust framework for neural network architecture-based false news identification. The fusion of Bi-LSTM and attention mechanisms holds promise in significantly impacting the fight against misinformation, fostering a more discerning society.

Looking ahead, future research will explore techniques such as Transfer Learning and Federated Learning to design more intelligent fake news detection systems. The SA-BiLSTM model, proposed here, will undergo testing in various domains, including disease, fraud, stock markets, and more, ensuring its applicability across diverse contexts.

ACKNOWLEDGMENT

The authors appreciate the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) for supporting and supervising this project.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

AVAILABILITY OF DATA AND MATERIAL

The whole ISOT dataset was selected from reliable, actual sources [12].

AUTHORS' CONTRIBUTIONS

All authors equally contributed in the article.

REFERENCES

[1] F. Farhangian, R. M. O. Cruz, and G. D. C. Cavalcanti, "Fake news detection: Taxonomy and comparative study," *Inf. Fusion*, vol. 103, Mar. 2024, Art. no. 102140.
[2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.
[4] T. Jiang, J. P. Li, A. U. Haq, A. Saboor, and A. Ali, "A novel stacking approach for accurate detection of fake news," *IEEE Access*, vol. 9, pp. 22626–22639, 2021.
[5] 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.
[6] A. Montieri, D. Ciuonzo, G. Aceto, and A. Pescapé, "Anonymity services tor, I2P, JonDonym: Classifying in the dark (Web)," *IEEE Trans. Dependable Secur. Comput.*, vol. 17, no. 3, pp. 662–675, May 2020.
[7] 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.
[8] S. Hakak, M. Alazab, S. Khan, T. R. Gadekallu, P. K. R. Maddikunta, and W. Z. Khan, "An ensemble machine learning approach through effective feature extraction to classify fake news," *Future Gener. Comput. Syst.*, vol. 117, pp. 47–58, Apr. 2021.

[9] 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.

[10] S. R. Sahoo and B. B. Gupta, "Multiple features based approach for automatic fake news detection on social networks using deep learning," *Appl. Soft Comput.*, vol. 100, Mar. 2021, Art. no. 106983.

[11] A. Choudhary and A. Arora, "Linguistic feature based learning model for fake news detection and classification," *Expert Syst. Appl.*, vol. 169, May 2021, Art. no. 114171.

[12] F. A. Ozbay and B. Alatas, "Fake news detection within online social media using supervised artificial intelligence algorithms," *Phys. A, Stat. Mech. Appl.*, vol. 540, Feb. 2020, Art. no. 123174.

[13] A. M. Braşoveanu and R. Andonie, "Semantic fake news detection: A machine learning perspective," in *Proc. 15th Int. Work-Conf. Artif. Neural Netw.*, Cham, Switzerland: Springer, 2019, pp. 656–667.

[14] T. Saikh, A. Anand, A. Ekbal, and P. Bhattacharyya, "A novel approach towards fake news detection: Deep learning augmented with textual entailment features," in *Proc. 24th Int. Conf. Appl. Natural Lang. Inf. Syst.*, 2019, pp. 345–358.

[15] A. Jain, A. Shakya, H. Khatter, and A. K. Gupta, "A smart system for fake news detection using machine learning," in *Proc. Int. Conf. Issues Challenges Intell. Comput. Techn. (ICICT)*, vol. 1, Sep. 2019, pp. 1–4.

[16] M. Samadi, M. Mousavian, and S. Momtazi, "Deep contextualized text representation and learning for fake news detection," *Inf. Process. Manage.*, vol. 58, no. 6, Nov. 2021, Art. no. 102723.

[17] I. K. Sastrawan, I. P. A. Bayupati, and D. M. S. Arsa, "Detection of fake news using deep learning CNN–RNN based methods," *ICT Exp.*, vol. 8, no. 3, pp. 396–408, Sep. 2022.

[18] A. Agarwal, M. Mittal, A. Pathak, and L. M. Goyal, "Fake news detection using a blend of neural networks: An application of deep learning," *Social Netw. Comput. Sci.*, vol. 1, no. 3, pp. 1–9, May 2020.

[19] Z. Chen, F. Zhuang, L. Liao, M. Jia, J. Li, and H. Huang, "A syntactic multi-level interaction network for rumor detection," *Neural Comput. Appl.*, vol. 36, no. 4, pp. 1713–1726, Feb. 2024.

[20] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.

[21] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[22] N. Parmar, A. Vaswani, J. Uszkoreit, L. Kaiser, N. Shazeer, A. Ku, and D. Tran, "Image transformer," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 4055–4064.

[23] A. Vaswani, N. Shazeer, and N. Parmar, "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[24] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.

[25] A. Mousa and B. Schuller, "Contextual bidirectional long short-term memory recurrent neural network language models: A generative approach to sentiment analysis," in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics*, vol. 1. Valencia, Spain: Association for Computational Linguistics, Apr. 2017, pp. 1023–1032.

[26] S. Khan, A. Kamal, M. Fazil, M. A. Alshara, V. K. Sejwal, R. M. Alotaibi, A. R. Baig, and S. Alqahtani, "HCovBi-caps: Hate speech detection using convolutional and bi-directional gated recurrent unit with capsule network," *IEEE Access*, vol. 10, pp. 7881–7894, 2022.

[27] X. Fang, H. Wu, J. Jing, Y. Meng, B. Yu, H. Yu, and H. Zhang, "NSEP: Early fake news detection via news semantic environment perception," *Inf. Process. Manage.*, vol. 61, no. 2, Mar. 2024, Art. no. 103594.

[28] D. Choudhury and T. Acharjee, "A novel approach to fake news detection in social networks using genetic algorithm applying machine learning classifiers," *Multimedia Tools Appl.*, vol. 82, no. 6, pp. 9029–9045, Mar. 2023.

[29] R. Sushila, R. Divya, and B. Rajesh, "Design and analysis of predictive model to detect fake news in online content," in *Artificial Intelligence, Blockchain, Computing and Security*. Boca Raton, FL, USA: CRC Press, 2023, pp. 102–106.

[30] A. Wani, I. Joshi, S. Khandve, V. Wagh, and R. Joshi, "Evaluating deep learning approaches for COVID19 fake news detection," in *Proc. Int. Workshop Combating Online Hostile Posts Regional Lang. During Emergency Situation*. Cham, Switzerland: Springer, 2021, pp. 153–163.

[31] D. K. Sharma and S. Garg, "IFND: A benchmark dataset for fake news detection," *Complex Intell. Syst.*, vol. 9, no. 3, pp. 2843–2863, Jun. 2023.

WANG JIAN was born in 1973. He received the master's degree from the University of Electronic Science and Technology of China, in 2006. He is currently a Professor with Neijiang Normal University, Neijiang, Sichuan, China. He is also an Associate Professor. His main research interests include operating systems and computer networks.

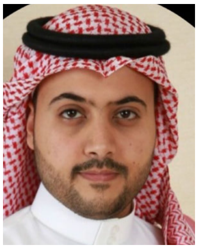
JIAN PING LI is currently the Chairperson of the Computer Science and Engineering College and the Model Software College, University of Electronic Science and Technology of China. He is also the Director of the International Centre for Wavelet Analysis and Its Applications. He holds roles, such as a Technical Adviser and a dozen academic and social positions at the National Science and Technology Award Evaluation Committee; the National Natural Science Foundation Committee of China; and the Ministry of Public Security of the People's Republic of China. He is the Chief Editor of the International Progress on Wavelet Active Media Technology and Information Processing. He is also an Associate Editor of the *International Journal of Wavelet Multimedia and Information Processing*.

MUHAMMAD ATIF AKBAR is currently pursuing the M.S. degree with Mohi-ud-Din Islamic University (MIU), Nerian Sharif, Azad Jammu and Kashmir, Pakistan. He is a Lecturer with the Government Boy's Degree College, Fatehpur Thakiala, Pakistan. His research interests include machine learning, deep learning, the IoT, web analytics, semantic web, concerned technologies, and algorithms.

AMIN UL HAQ received the Ph.D. degree in computer science (software engineering) from the University of Electronic Science and Technology of China. He is an Experienced Researcher with more than 12 years of cutting-edge research and teaching experience in prestigious institutes, including the National University of modern languages (NUML), Abdul Wali Khan University Mardan, Bunar campus, and Agricultural University, Peshawar, Pakistan. He is Postdoc fellow with the School of Computer Science and Engineering, University of Electronic Science and Technology of China. He has first and co-authored more than 88 research papers, including 51 in leading international journals SCI of rank JCR 1, 2, 3, 4, and 5 papers are in JCR rank 1 and Q1 SCI journals(namely Knowledge-based Systems, Expert Systems with Applications, IEEE Journal of Biomedical and Health Informatics, Computer and Electrical Engineering, Scientific Reports, *Journal of Ambient Intelligence and Humanized Computing*, IEEE Access, and Sensors) and 36 in peer-reviewed international conference proceedings and one Patent. During his career, he has taught various courses at the Undergraduate (UG) and Postgraduate (PG) levels. He has more than 2800 Google citations, an h-index of 24, an I-10 index of 43, and an impact factor of more than 160.40. Also, sponsor the IDITR IEEE conference. He is an invited reviewer for numerous world-leading high-impact journals (reviewed 70+ journal papers to date). His area of research is medical data analysis to diagnose diseases using Machine learning, Deep learning models, Transfer Learning, and Federated learning Techniques. He regularly organizes timely special sessions and workshops for several Flagship IEEE conferences and is an Active Member of the IEEE International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP).

SHAKIR KHAN received the Ph.D. degree in computer science. He is currently an Associate Professor with the College of Computer and Information Sciences (CCIS), Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia. He has more than 15 years of rich national and international teaching, research, and IT experience, including five years with King Saud University, Riyadh, for research, training, and teaching. He is delivering these projects timely. He is also doing collaborative research with the University of Electronic Science and Technology China (UESTC) and other professors globally. He has sound research knowledge along with the supervision of more than a dozen master's scholars in different fields of computer science. He has published more than 70 publications in reputed journals and conferences. He has also published two books and three patents. His research interests include data mining, data science, AI, cloud computing, the IoT, big data and analytics, bioinformatics, e-learning, machine learning, and deep learning. He contributed to many international journals and conferences as an editor, a member of the advisory board, a reviewer, a program committee member, and a keynote speaker. He has been acknowledged by Imam University and received the Research Excellence Award, in 2019, and received as the Leader of the Programming Team, Prince Sultan University, Riyadh, as the team was the third winner out of 40 national teams from different Saudi universities. He has been awarded three research grants from the Deanship of Scientific Research, Imam University, and one from the Ministry of Education, Saudi Arabia. One grant is awarded as a PI for the International Research Partnership Program with one of the computer science professors at Aligarh Muslim University (AMU), India.

REEMIAH MUNEER ALOTAIBI received the B.Sc. degree from Imam University, and the M.Sc. and Ph.D. degrees in information management from Leeds Beckett University, U.K. Moreover, she has some courses, such as the Professional Development Diploma in project management with the Midlands Academy of Business and Technology, U.K. She is currently an Assistant Professor with the College of Computer Sciences and Information, Imam Muhammad Ibn Saud Islamic University (IMSIU), Saudi Arabia. She has also authored papers published in scientific journals, book chapters, and conference materials. Her research interests include knowledge management, disaster management, e-government, data analysis and engineering, information seeking, and technology acceptance and adoption. She has acted as a Reviewer of *International Journal of Civic Engagement and Social Change* (IGI Global).



SAAD ABDULLAH ALAJLAN received the M.S. degree in information management from Syracuse University, in 2016, and the Ph.D. degree in computer science from the University of Liverpool, in 2021, for his dissertation on generating an RDF dataset from Twitter data. He is currently an Assistant Professor of computer science with Imam Mohammad Ibn Saud Islamic University. His research interests include artificial intelligence (deep learning) in several domains, such as healthcare.