Iraqi Journal for Computer Science and Mathematics

Journal Homepage: http://journal.esj.edu.iq/index.php/IJCM
e-ISSN: 2788-7421 p-ISSN: 2958-0544

Detecting Arabic Misinformation Using an Attention Mechanism-Based Model

Mohammed Haqqi Al-Tai[1], Bashar M. Nema[1,*]

[1]Department of Computer Science, Mustansiriyah University, Baghdad, IRAQ

*Corresponding Author: Bashar M. Nema

DOI: https://doi.org/10.52866/ijcsm.2024.05.01.020
Received Augest 2023 ; Accepted November 2023 ; Available online February 2024

ABSTRACT: The proliferation of fake news or misinformation, commonly referred to as fake news, has a significant effect on a global scale, as it is aimed at influencing public opinion as well as crowd decision-making. It is therefore crucial to verify the truthfulness of news before it is released to the public. Today, most studies on early detection of Arabic misinformation rely on machine learning methods and transformer-based models. Therefore, in the current study, we used deep learning techniques to propose a model for detecting Arabic misinformation by leveraging the contextual features of news article content. The proposed model was built based on BiLSTM and the attention mechanism. To extract features from Arabic text, we utilized a pre-trained AraBERT model, which generates contextual embeddings from text, then are fed to the BiLSTM layer as input features. Moreover, we investigated the effectiveness of the attention mechanism in improving the overall performance of the model by configuring model architecture to exclude the attention mechanism and comparing the results. Two datasets were utilized to train and evaluate the proposed model, namely, the AraNews and ArCovid19-Rumors datasets. Experimental results showed that the proposed model outperformed other existing models, achieving an accuracy of 0.96 on the ArCovid19-Rumors dataset and 0.90 on the AraNews dataset. This remarkable performance was due to the capability of the attention mechanism to enhance the overall performance, allowing the model to capture the relationship between textual features.

Keywords: Misinformation, Fake news, Attention mechanism, BiLSTM, LSTM, Bert, AraBert, Contextual embeddings.

1. INTRODUCTION

Misinformation represents a serious problem that has a strong impact and pernicious consequences on society. While misinformation is not a new phenomenon, the technological revolution, especially in the social communication field, has contributed to accelerating the spread of misinformation as well as to increasing its harm. According to the World Health Organization (WHO), four studies reviewed misinformation on social media during the COVID-19 pandemic and found that 51% of posts correlated to vaccines increased up to 28.8% of posts about COVID-19, and increased up to 60% of posts correlated to pandemics [1].

In the same context, the Pew Research Center [2] reported in 2016 that approximately 62% of American adults obtain news from social media. Nowadays, online social networks have become a significant resource for accessing and sharing news, and various actors like people, companies, and institutions utilize it. Each kind has different motivations, like economic benefits, entertainment, or causing harm to people. A massive part of this news spread on online social networks is considered unreliable, and its source cannot be ascertained.

According to a survey conducted by the Knight Foundation, Americans estimate that 65% of the news they encounter on social media is false. Moreover, in social networks, misinformation typically spreads more rapidly, deeply, and broadly [3]. The widespread misinformation is not only limited to misleading people and making them accept fake facts and change their reaction to the real news but also causes a state of confusion to the reader, which impedes their ability to distinguish between falsehood and truth, also breaking the trust of the information ecosystem [4]. Political misinformation can affect public opinion and undermine the democratic system, resulting in dissatisfaction and even violence among the general public. One example is the case in the 2016 U.S. presidential elections where misinformation intentionally spread to affect public opinion. The amount of misinformation disseminated in relation to the presidential election quickly surpassed the number of articles published by credible news sources [5]. The propagation of misinformation or unsubstantiated rumors can have far-reaching negative

consequences, including economic harm and widespread panic. A study on misinformation in financial found that trading activity increased by more than 50% with the impact of misinformation. Also, social media bots spread financial misinformation to manipulate the stock market [6]. The misinformation is consistently given as factually accurate, although it is not.

In today's society, we believe what we see on social media without attempting to verify the credibility of the information. Since people frequently lack the time to check sources and the veracity of news, the detection of misinformation is essential, and this area has received significant interest from the research community. To handle misinformation, numerous aspects should be considered, such as early detection of misinformation in newly produced information to prevent the chances of widespread propagation with potentially harmful effects. Moreover, detecting misinformation sources helps identify malicious actors who initiate and spread misinformation [7].

While there is a significant body of research dedicated to identifying instances of fake news in the English language, the detection of Arabic fake news remains a relatively underdeveloped area of inquiry. This can be attributed to the scarcity of accessible datasets and the complexities associated with addressing the obstacles inherent in the Arabic language. Most previous studies on Arabic misinformation detection were presented based on traditional machine learning algorithms and language models based on transformer architecture. Therefore, the current study aimed to detect Arabic misinformation using deep learning techniques. The proposed model in this study utilized the AraBERT model to extract relevant features and to capture long-range dependencies from the input sequence through the BiLSTM architecture. The attention mechanism was then implemented as a layer in the proposed model, following the BiLSTM layer, to focus on the most relevant parts of the BiLSTM layer's outputs. Adding an attention mechanism to the model improved the overall performance of the model. In brief, our contributions include the following:

1. Proposing a deep learning model based on bidirectional LSTM and attention mechanism for detecting Arabic misinformation.
2. Proving that deep learning-based models outperform traditional machine learning models in Arabic misinformation detection task.
3. Improving the accuracy of detection of Arabic misinformation using an attention mechanism, as compared to other existing studies that utilized the AraNews and the ArCovid19-Rumors datasets.

2.RELATED WORKS

In recent years, the spread of misinformation has become a major concern not only in English but also in numerous other languages, including Arabic. However, due to the distinct linguistic and cultural characteristics of the Arabic language, detecting misinformation in the language presents unique challenges. This section explores prior research and approaches related to the detection of Arabic misinformation. In [8], the author presents a machine learning model that aims to identify fake Arabic news by utilizing content- and context-related features. The model developed by the author performs classification on a small dataset of 800 tweets related to Arabic political news. This dataset was human-labeled and collected from the social media platform, Twitter. Several contextual features such as demographic information associated with the account, the account name, and the number of followers are incorporated as non-linguistic features. Furthermore, various machine learning algorithms such as Naive Bayesian (NB), Support Vector Machine (SVM), and Decision Tree (DT), were tested. The J48 DT classifier yielded the highest predictive accuracy, 89.9%. In [9], the author proposes multiple machine learning models to classify satirical fake Arabic news by exploiting linguistic features inherent in Arabic satirical fake news. In addition, the author introduces a novel dataset comprised of 3,185 fake news articles and 3,710 real news articles. The experimental results demonstrate that Convolutional Neural Networks (CNNs) with fastText word embedding can achieve substantial classification accuracy, reaching as high as 98.6%. In comparison, Naive Bayes Multinomial NB with both Bag-Of-Words (BOW) and TF-IDF as well as XGBoost with TF-IDF values achieve accuracies of 96.23% and 96.81%, respectively. In [10], the authors mention the lack of data for detecting fake news in Arabic. They propose to focus their research on the readily available true stories on the internet, utilizing a part of speech tagger (POS). In their study, various BERT-based models (mBert, XLMR, and Arabert) were employed, with Arabert demonstrating particularly strong performance. The evaluation of their models was based on accuracy, and it yielded results of 79.39%, 82.77%, and 89.25%, respectively. In [11], the authors present a comprehensive investigation examining various deep neural network models and transformer architecture-based models for the detection of Arabic fake news on social media platforms. The models were trained on three datasets, including ArCOV19-Rumors, AraNews, and Ans; they were evaluated using the COVID-19-Fakes dataset .The experimental results demonstrate that transformer-based models outperformed deep neural network in identifying fake and authentic news. On the ArCOV19-Rumors dataset, QARiB and ArBERT demonstrate a high level of performance, obtaining an accuracy of up to 0.95%, whereas on the AraNews dataset, QARiB and ArBERT achieved a lower level of accuracy.

In [12], the author employs the AraNews dataset as a foundation for model development. The methodology included the implementation of the Term Frequency-Inverse Document Frequency (TF-IDF) technique, a strategic approach used for feature extraction in the form of word vectors. Subsequent steps involved the utilization of Machine Learning techniques, namely, the Random Forest Classifier (RF), Naive Bayes (NB), and Logistic Regression (LR), to

predict Arabic fake news. Among the three models, the RF exhibited superior performance, with an accuracy of 0.866%. In comparison, the NB and LR algorithms yielded slightly lower accuracies, 0.844% and 0.859%, respectively.

In [13], the authors propose a model architecture for detecting Arabic fake news based on only textual features. Three datasets are utilized to evaluate the performance of eight distinct machine learning algorithms. Moreover, they conducted experiments using five separate combinations of deep learning algorithms, including CNN and LSTM. The outcomes show that the BiLSTM outperformed other deep learning methods, achieving an accuracy rate of 77% on the dataset of size 4561.

In [14], the authors introduce deep learning-based models for identifying fake news within Arabic tweets by exploiting the content of the news in tweets. They conducted experiments using CNN and BiLSTM architecture and investigated the utilization of five distinct word embeddings, including word2vec, FastText, and a BERT-based model for extracting textual features. The outcomes of these experiments indicated that the MARBERT-CNN architecture yielded remarkable results, achieving an accuracy score of 0.86% on the ArCOV19-Rumors dataset.

Table 1. – Summary of related works

Ref	Year	Proposed Methods	Dataset	Features	Accuracy
[8]	2018	Naïve Bayesian (NB), Support Vector Machine (SVM), Decision Tree (DT),	Arabic political news dataset	contextual features	Naïve Bayesian 78%, Support Vector Machine 80%,Decision Tree 89.9%
[9]	2020	CNN, Naïve Bayes, XGBoost	novel dataset	linguistic features	CNN 98.6, Naïve Bayes 96.23%, XGBoost 96.81%,
[10]	2020	mBert, XLMR, AraBert	AraNews	contextual features	mBert 79.39%, XLMR 82.77%, AraBert 87.21%
[11]	2021	CNN, RNN, GRU, AraBERTv02, AraGPT2, QARiB, ArBert	ArCOV19-Rumors, AraNews, Ans, COVID-19-Fakes	contextual features	Highest accuracy achieved with ArBert 95% on A Covid 19-rumors and with QARiB 80% on AraNews.
[12]	2022	Random Forest Classifier (RF), Naive Bayes (NB), Logistic Regression (LR)	AraNews	contextual features	Random Forest 0.866%, Naive Bayes 0.844% , Logistic Regression 0.859%,
[13]	2022	CNN, LSTM, BiLSTM, CNN + LSTM, CNN + BiLSTM	Merged dataset	textual features	CNN 0.71%,LSTM 0.76%, BiLSTM 0.77%, CNN + LSTM 0.76%, CNN + BiLSTM 0.75%
[14]	2023	BiLSTM, MARBERT-CNN	ArCOV19-Rumors	textual features	BiLSTM 80% , MARBERT-CNN 86%

3. METHODOLOGY

This paper focuses on detecting Arabic misinformation, with the aim of detecting the truthfulness or falsehood of news. A review of the literature shows that multiple machine learning-based approaches have been proposed for misinformation detection, although these models are lacking in terms of high performance. We try in this research to enhance performance by utilizing an attention mechanism-based deep learning model.

This section details the proposed model's architecture. It also includes details about the datasets used to train and evaluate models as well as a brief background of the pre-trained BERT model, attention mechanism, BiLSTM architectures, the dataset used, and the preprocessing techniques. Fig (1) illustrates the proposed methodology for content-based misinformation classification.

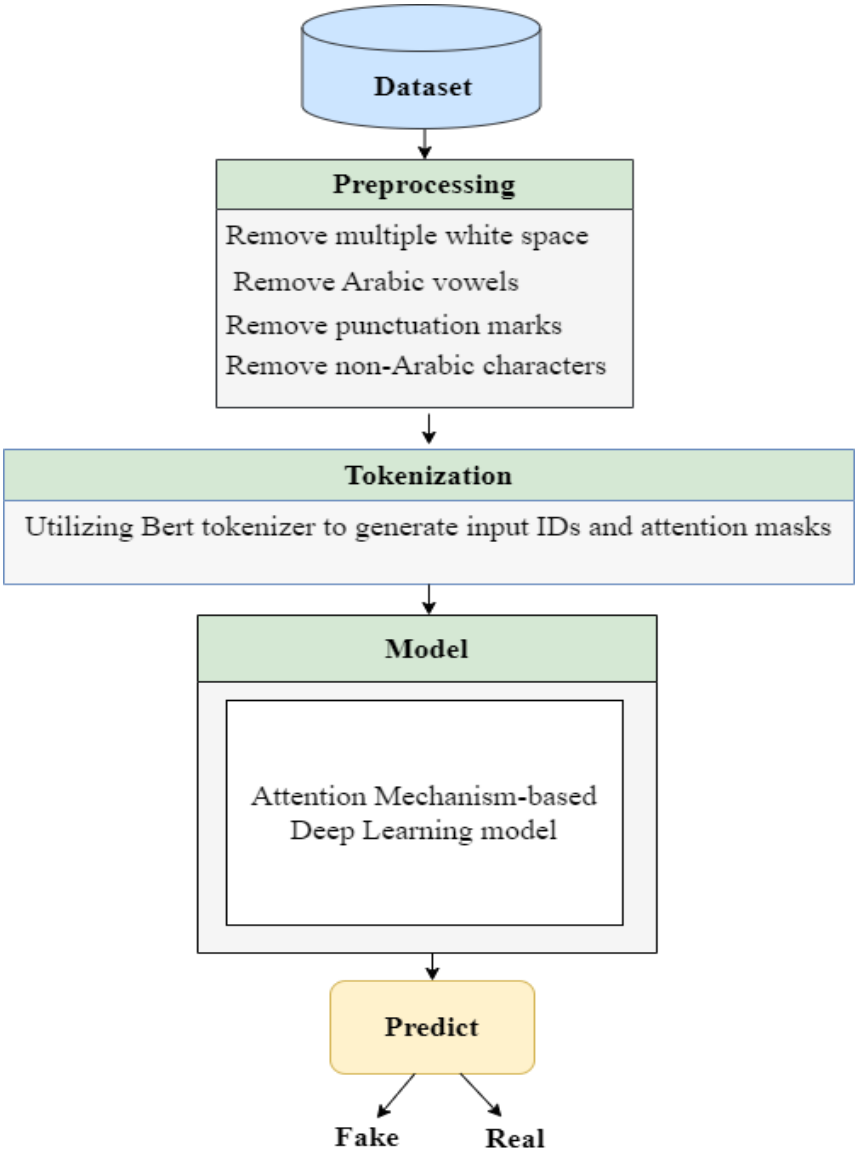


FIGURE 1. - Proposed Methodology for Misinformation Detection

3.1 PREPROCESSING

Preprocessing the input data is a crucial step before applying a deep learning model. Eliminating noise from the dataset can significantly improve the performance of a neural network. However, when it comes to Arabic language, which is an orthographic language that relies on the word's form, applying preprocessing techniques becomes challenging without altering the meaning of words [15, 16]. In our work, preprocessing starts with cleansing data in datasets by eliminating irrelevant data that are worthless for misinformation detection and that could be considered noisy. This process involves several essential steps:

- Removed punctuation marks
- Removed non-Arabic characters
- Removed Arabic vowels such as diacritics and teshkeel
- Removed Arabic tatweel character
- Removed trailing whitespaces.

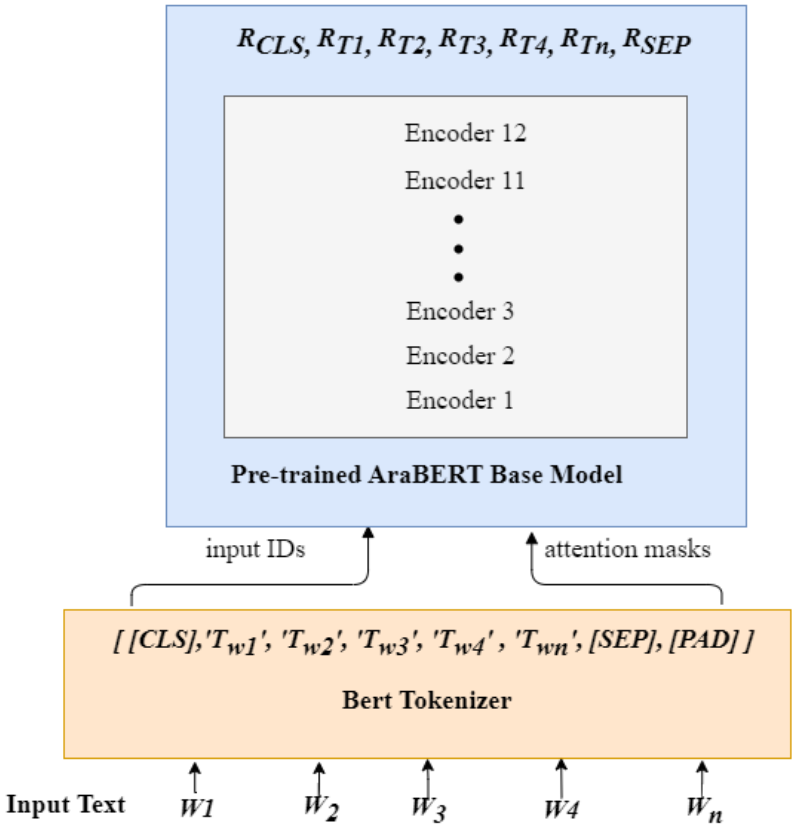


FIGURE 2. - AraBERT model for generating contextual embeddings.

3.2 Tokenization

After preprocessing the text contents of the datasets, each text of a news article is tokenized using the BERT Tokenizer to obtain input IDs and attention masks. Then, both the input IDs and attention masks are fed to the BERT model to generate contextual embeddings, which are then used as input features for the neural network, as we will discussion the process in details in next section.

3.3 BERT

BERT, which stands for Bidirectional Encoder Representations from Transformers, is constructed from a transformer attention mechanism [17]. This mechanism is capable of understanding the contextual relationships between words. The transformer is comprised of two elements: an encoder responsible for reading textual input and a decoder that is responsible for making task-based predictions. Unlike directional models that process the text input in a sequential manner, the transformer encoder reads all words at once, which gives it a bidirectional property. As a result, the model is able to learn the contextual meaning of a word from all the words that surround it. There exist multiple variants of pre-trained BERT models [18], and Table (2) shows the two models most frequently utilized.

Table 2. - Details of the two most commonly used BERT models [17]

Model	Encoder stack layer	Multi-head Attention head	Hidden units	Parameters
Pre-training Bert Base model	12	12	768	110M
Pre-training Bert Large model	24	24	1024	340M

The input data is transformed into an understandable format before being passed to the pre-trained BERT model. This transformation involves passing the preprocessed input data through the BERT Tokenizer to produce input IDs and attention masks, which are then used as input into the BERT model to get contextual embeddings (vector representation), as depicted in Fig (2). BERT Tokenizer divides each input text into individual words or sub-words to obtain the tokens. After the text is tokenized, special tokens indicating the beginning ([CLS]) and end ([SEP]) of the

text are added. Each token is mapped to a unique integer ID based on BERT's vocabulary, which includes all the tokens that the BERT model was trained on. These IDs form the input IDs, a list of integers representing the tokens in the text. Also, since the dataset contains texts of varied lengths, it is necessary to ensure that all articles have the same length. This process is achieved through padding, where shorter text is padded with extra tokens ([PAD]) to match the length of the longest text. Since these tokens ([PAD]) do not contain useful information, the BERT model needs to distinguish between the padding tokens and the real tokens from the input text; therefore, the attention mask assigns a value of 1 to real tokens and 0 to ([PAD]) tokens to instructs the BERT model to ignore the padding tokens during attention score calculations.

In this work, we utilized AraBERT, an Arabic pre-trained model based on the BERT architecture [19], to extract feature representations from Arabic texts that effectively capture the inherent semantic and contextual information. We also extracted contextual embeddings of tokens from the last hidden layer, which are then used as input features in our model.

3.4 LSTM

The LSTM (long short-term memory) network is an advanced version of the recurrent neural network (RNN) that addresses a vanishing gradient problem by utilizing a specialized hidden layer unit called memory cells [20]. These memory cells have self-connections that preserve the time state of the network, which is managed through three primary gates: the input, output, and forget. The input and output gates regulate the flow of data inputs and outputs of the memory cell into the rest of the neural network, while the forget gate is a unique addition to the memory cell that allows the passage of high-weighted output information from one neuron to the subsequent one. The data stored within the memory depends on the results of high activation; if the input unit exhibits high activation, the information is retained in the memory cell. Moreover, if the output unit shows high activation, it transmits the information to the following neuron. Otherwise, input data with significant weights remains in the memory cell [21].

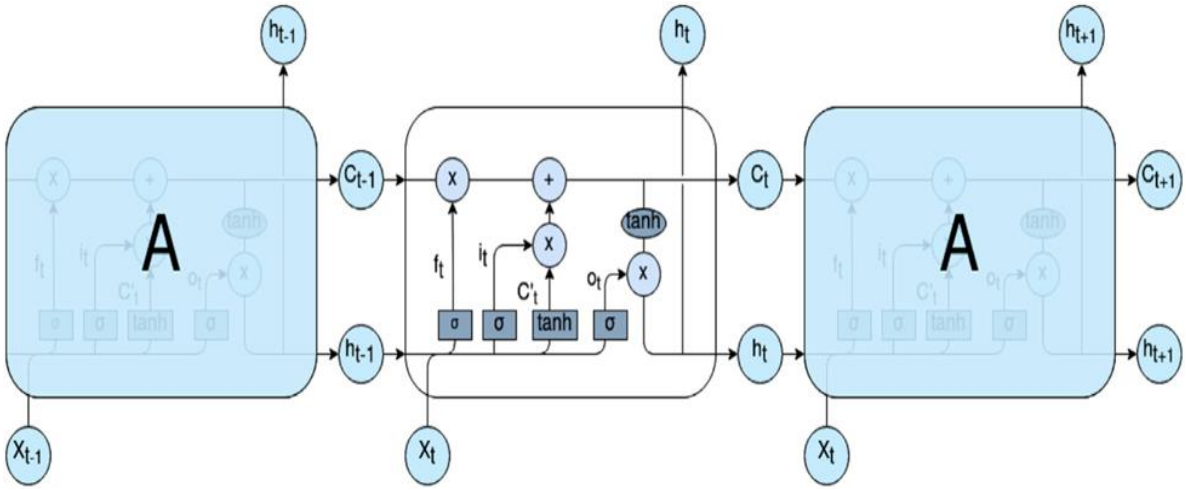


FIGURE 3. - LSTM architecture [22]

In comparison to traditional RNNs, LSTM can handle long-term dependencies and overcome the vanishing gradient problem, which can arise when gradients become too small for network training. LSTM computing is performed by the following equations:

$$i_{nput}\, g_{ate} = sigmoid\,(W_i \,.\, [h_{t-1}\,, x_t\,] + b_i\,) \quad (1)$$

$$f_{orgat}\, g_{ate} = sigmoid\,(W_f \,.\, [h_{t-1}\,, x_t\,] + b_f\,) \quad (2)$$

$$o_{utput}\, g_{ate} = sigmoid\,(W_o \,.\, [h_{t-1}\,, x_t\,] + b_o) \quad (3)$$

$$g = tanh\,(W_c \,.\, [h_{t-1}, x_t] + b_c\,) \quad (4)$$

$$c_t = (f * c_{t-1}) + (g * i) \quad (5)$$

$$h_t = tanh\,(c_t) * o_t \quad (6)$$

In Equations 1 through 6, W_i , W_f , W_o , W_c , and b_i , b_f , b_o , b_C denote the weight and bias variables for the three gates and the memory cell respectively. Here, h_{t-1} represents the preceding hidden layer units, which are added to the weights of the three gates. In eq.4, g is the candidate state value. Upon processing Eq. 5, C_t becomes the current memory cell unit. Eq.6 illustrates the element-wise multiplication of the outputs from the previous hidden unit and the previous memory cell unit. Additionally, non-linearity is introduced to the three gates via *tanh* and *sigmoid* activation functions, as depicted in Eq. 1 through 6. In this context, $t - 1$ and t represent the previous and current time steps, respectively.

We used BiLSTM, which is a type of neural network architecture that allows the model to take into account both past and future contexts when making predictions about a given sequence of data. BiLSTM is composed of two LSTM layers that are capable of processing input sequences in a bidirectional manner, namely, forward and backward. This allows the model to capture information from both past and future contexts. Forward LSTM processes the data from the start of the sequence to the end (left to right), whereas backward LSTM processes data from the end of the sequence to the start (right to left). The hidden layer outputs of each LSTM layer in the forward ($h^{\rightarrow}$) and backward ($h^{\leftarrow}$) directions are computed through Eq. (1) to (6). Then $h^{\rightarrow}$ and $h^{\leftarrow}$ are concatenated to form the final output sequence of the BiLSTM (h_t).

$$h_t = [h_t^{\rightarrow}, h_t^{\leftarrow}] \qquad (7)$$

The output of BiLSTM is a sequence of hidden layers. This output sequence contains information from both past and future contexts, which can be beneficial for tasks such as sentiment analysis, machine translation, and text classification.

3.5 ATTENTION MECHANISM

Bahdanau et al. [23] were the first to propose the attention mechanism in 2015 as a solution to the information bottleneck problem in RNN-based models such as GRUs and LSTMs for machine translation (MT). These models process input in a recurrent fashion and the decoder part struggles to extract results because it only receives an intermediate representation (context vector) as input. Bahdanau proposed an attention mechanism that uses weights on intermediate hidden values to align the amount of attention the model must pay to the input in each decoding step. This mechanism helps the model to focus on important parts of the input, and has been used in various deep learning architectures, where it aids deep learning models to effectively capture relevant information [24].

Neural network-based classification systems traditionally model data as numerical vectors consisting of low-level features. However, in models that utilize attention mechanisms, not all features are given equal importance. The attention mechanism assigns different weights to different features depending on their importance for the task at hand, thereby enabling the model to conceptualize the data more effectively. Further, the attention mechanism equips the neural model with the ability to rank features according to their relevance. The principal concept of this mechanism is to calculate a weight distribution over the input features, assigning higher values to those features deemed more important. The attention layer is composed of alignment scores, attention weights, and context vector [25], which are computed using the following equivalents:

$$h_{it} = tanh(\mathrm{W}.h + b) \quad (8)$$

$$a_{it} = \frac{exp(h_i^T h_w)}{\Sigma_i exp(h_i^T h_\omega)} \quad (9)$$

$$s_i = \sum_i \alpha_i h_i \qquad (10)$$

Here, h_{it} is the outcome of a full connection operation (with *tanh* activation) applied to the hidden state vector h_i; the weight matrix and bias for this operation are symbolized by W and b, respectively. h_w is the *alignment* vector, initialized randomly and updated throughout the training process. α_{it} is the attention score (or weight) for the i-th word in the sequence. This is calculated by applying a *softmax* function to the scores, so that they form a valid probability distribution. Finally, the context vector, si, is a weighted sum of the hidden states, where the weights are the calculated attention scores. This context vector represents a weighted combination of the input that highlights those elements deemed more important by the attention mechanism.

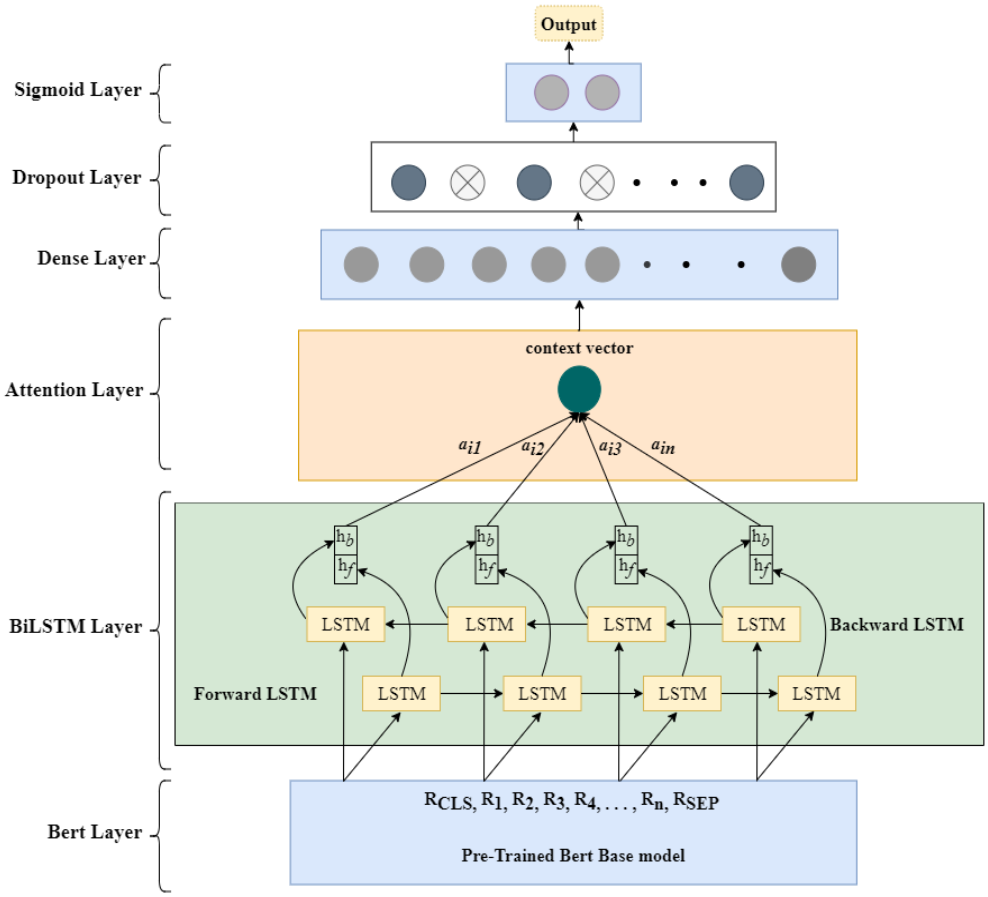


FIGURE 4. - Proposed model architecture

3.6 PROPOSED MODEL ARCHITECTURE

We propose a deep learning model based on an attention mechanism for detecting Arabic misinformation. The architecture of this model is presented in Fig 4. In proposed neural network, the first layer, the BERT layer, serves as a feature extractor. It extracts contextual features from the content of news articles using the AraBERT (Bert-base-arabertv02) model, which has 12 encoders and 768 hidden units.

The AraBERT model takes inputs in the form of input IDs and attention masks derived from the input text. This task is performed by the BERT tokenizer, which generates the input IDs and attention masks. These are then fed into the AraBert model, which generates a vector representation for each token, each with a hidden size of 768. AraBERT provides context-dependent representations that enhance the capability of BiLSTM to comprehend word semantics more effectively. The outputs of the Bert layer are fed to the BiLSTM layer as input features. BiLSTM could capture past and future contextual information by applying forward LSTM and backward LSTM to these features, which the information it obtains can be considered two different textual representations. Sequences of hidden states of the BiLSTM layer are then passed as input to the attention layer, which generates a context vector corresponding to the learned input vectors. This context vector is then fed into a first dense layer. Following this is a dropout layer that prevents overfitting problems. The output is then fed into a second dense layer (output layer) to classify the input news article as fake or real. The output layer has one hidden neuron and uses the sigmoid activation function suitable for a binary classification task.

4. EXPERIMENTS AND RESULTS

4.1 DATASET

In this study, we utilized two datasets—namely, the ArCovid19-rumors dataset [26] and the AraNews dataset [10]—to train our model for detecting Arabic misinformation. ArCovid19-rumors is an Arabic COVID-19 Twitter dataset specifically designed for detecting Arabic misinformation. It comprises 3,584 tweets, with 1,753 as false news and 1,831 as true news. The AraNews dataset, which encompasses a vast collection of Arabic fake news, was compiled from numerous newspapers covering a wide range of topics. This dataset was collected from 15 Arabic countries, in addition to the United Kingdom and the United States of America. It consists of 108194 news articles labelled into two

classes, fake or not fake. We split each dataset into 90% for training and 10% for testing, and 10% of the training set is allocated to validation during training.

Table 3. - Analysis of the distribution of AraNews and ArCOV19-Rumors datasets into three sets

Dataset	Data	Fake	Real	Total
AraNews	Training set	45005	42632	87637
	Testing set	5555	5265	10820
	Validation set	4986	4751	9737
	All	55546	52648	108194
ArCOV19-Rumors	Training set	1429	1474	2903
	Testing set	176	183	359
	Validation set	148	174	322
	All	1753	1831	3584

4.2 EXPERIMENT SETUP

We implemented the proposed model on the CPU using Visual Studio Code software and utilized various Python libraries such as TensorFlow, Scikit-learn, and PyArabic. In addition, a pre-trained Arabic BERT model with embedding dimensions of 768 was utilized for the ArCOV19-Rumors dataset and AraNews datasets. The hyperparameters of our model were chosen by utilizing the Keras Tuner with Bayesian optimization technique that helped us to fine-tune the model's hyperparameters and to avoid overfitting issues. These hyperparameters were selected through a Bayesian optimization technique, taking into account search space for hyperparameters as in Table (4).

Table 4. - Search Space for Hyperparameter Ranges in Experiments Setup

Hyperparameter	Minimum Value	Maximum Value	Increase by the Step
BiLSTM Layer units	128	1024	32
Attention layer units	64	1024	32
Dense layer unit	64	1024	32
Dropout Rate	0.05	0.5	0.05
Learning rate	1e-5	1e-2	'Log'

Table 5. - Hyperparameters of our model with AraNews dataset

Hyperparameters	Value
BiLSTM hidden Units	576
Attention layer units	640
Dense layer unit	416
Dropout Rate	0.1
Learning rate	1e-3

Table 6. - Hyperparameters of our model with ArCovid19-Rumars dataset

Hyperparameters	Value
BiLSTM hidden Units	192
Attention layer units	512
Dense layer unit	256
Dropout Rate	0.25
Learning rate	1e-3

The hyperparameters in Tables (5) and (6) represent the best hyperparameters found through the Keras Tuner. We conducted experiments using the hyperparameters in Table (5) with the AraNwes Dataset and Table (6) with the ArCOV19-Rumors dataset. Moreover, the binary cross-entropy loss function and Adam optimizer were utilized to train our model for 10 epochs; also, early stopping with patience equal to 5 was used to automatically stop the training if the validation loss dropped five times continuously.

4.3 RESULTS AND DISCUSSION

To evaluate the performance of the proposed model, the experiment was conducted using both an ArCovid19-Rumors and the AraNews datasets. Various performance evaluation parameters—namely, accuracy, precision, recall, and F1-score—were employed to evaluate the model's performance.

$$\text{Accuracy} = \frac{Number\ of\ instances\ correctly\ Predict}{Total\ Number\ of\ instances} \quad (10)$$

$$\text{Precision} = \frac{True\ Positives}{(True\ Positives + False\ Positive)} \quad (11)$$

$$\text{Recall} = \frac{True\ Positives}{(True\ Positives + False\ Negatives)} \quad (12)$$

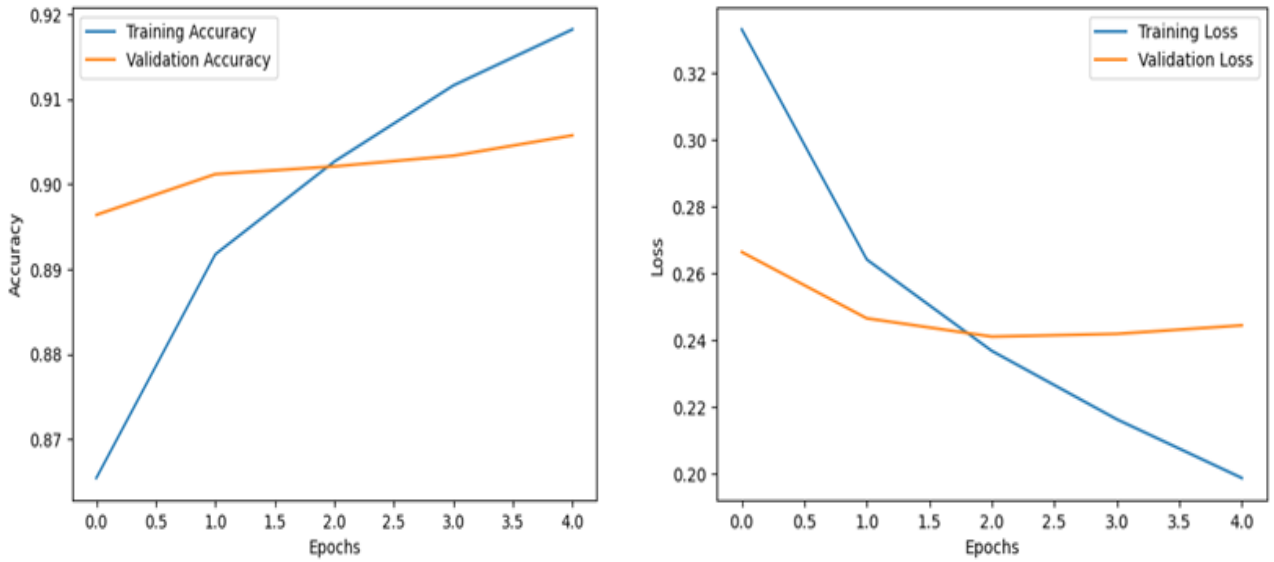
$$\text{F1 Score} = 2 * \frac{(\text{precision} * \text{recall})}{(\text{precision} + \text{recall})} \quad (13)$$

In investigating the capability of the attention mechanism to improve the overall performance of the model, we excluded the attention layer from the proposed model and kept the rest of the model structure to obtain two models, BiLSTM and Att-BiLSTM (BiLSTM with an attention mechanism). These models share an identical architectural blueprint, as delineated in Fig (4). However, the key distinction between them lies in excluding the attention layer from BiLSTM, while att-BiLSTM includes the attention layer. The experiment results obtained from training BiLSTM and Att-BiLSTM on the AraNews dataset and ArCOV19-Rumors dataset have shown that the Att-BiLSTM model yielded the highest performance metrics on both datasets, with 90% accuracy on AraNews and 96% on ArCOV19-Rumors. In contrast, the BiLSTM model did not perform as effectively as the att-BiLSTM model, which utilizes an attention mechanism to enhance its performance. The Att-BiLSTM model incorporates an attention mechanism that enables it to focus on specific parts of the input sequence, allowing it to capture relevant information. As a result, the BiLSTM model yielded lower performance scores, achieving an accuracy of 88% on AraNews and 92% on ArCOV19-Rumors.

Table 7. - Performance Comparison of Proposed Models on AraNews and ArCOV19-Rumors Datasets.

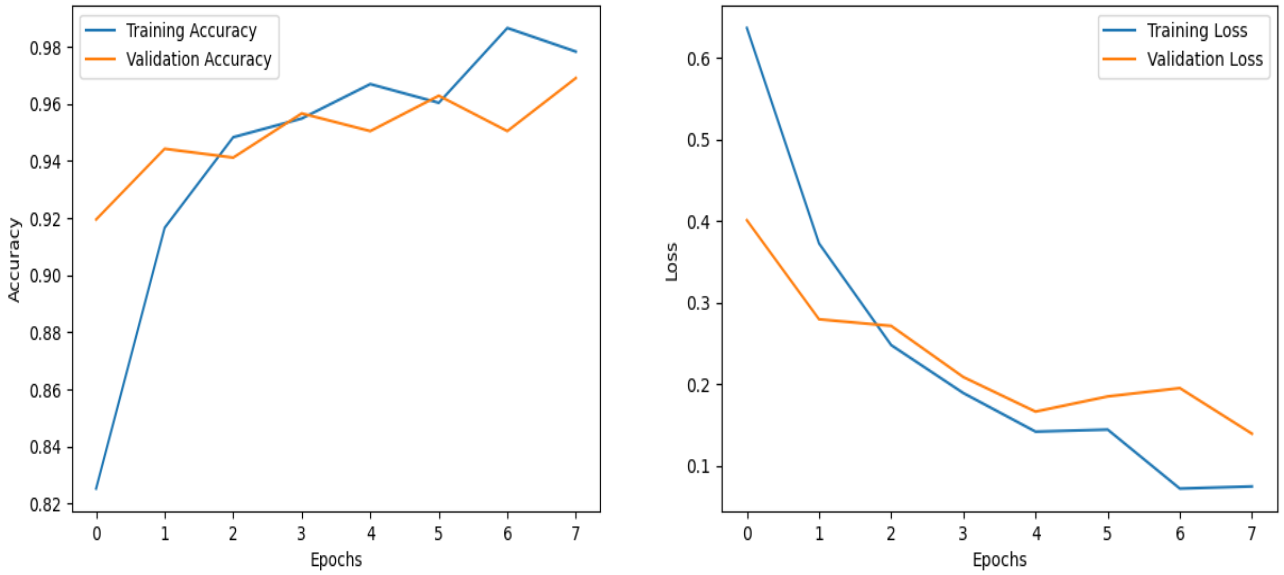
Dataset	Model	Accuracy	Precision	Recall	F1-Score
AraNews	BiLSTM	0.881	0.838	0.937	0.885
	Att-BiLSTM	**0.905**	0.861	0.960	0.907
ArCOV19-Rumors	BiLSTM	0.930	0.920	0.934	0.931
	Att-BiLSTM	**0.969**	0.982	0.960	0.970

Based on the results in Table (7) on the AraNews dataset, the BiLSTM and att-BiLSTM models yielded high recall scores of 93% and 96%, respectively, with the Att-BiLSTM demonstrating a 2.3% increase. The BiLSTM model had a precision score of 83%, while the Att-BiLSTM obtained the best result of 86%. Finally, the BiLSTM and att-BiLSTM models yielded 88% and 90% on the F1- score. In contrast, on the ArCOV19-Rumors dataset, the Att-BiLSTM model significantly outperformed the BiLSTM model in all performance metrics, achieving 98%, 96%, and 97% in terms of precision, recall, and the F1-score, respectively. Although the BiLSTM model achieved good results, with precision, recall, and F1-score of 92%, 93%, and 93%, respectively, its performance was lower than that of the Att-BiLSTM model, thus proving the efficacy of the attention mechanism in enhancing the overall model performance.



(a) (b)

FIGURE 5. - depicts the fluctuations of (a) accuracy and (b) loss of proposed model with AraNews dataset



(a) (b)

FIGURE 6. - depicts the fluctuations of (a) accuracy and (b) loss of proposed model with ArCovid19 dataset

To evaluate the performance of our model, we compared the performance of our proposed model with the baseline models in studies [10, 11, 12, 14] which utilized the same dataset. We compared our model with baseline models in four studies from the related works section that utilized the AraNews dataset and the ArCOV19-Rumors dataset, including:

- Nagoudi and El Moatez Billah utilized three models based on transformer architecture, with the highest accuracy of 90% obtained with AraBert when using the AraNews dataset.

- Al-Yahya et al. analyzed the utilization of neural networks and models based on transformer architecture. They obtained the highest accuracy up to 95% with both QARiB and ArBert on ArCOV19-Rumors, while QARiB and AraBert both achieved 80% accuracy on AraNews.
- Aljwari and Fatima examined several traditional machine-learning algorithms with the AraNews dataset, and the highest accuracy achieved 86% obtained by utilizing the Random Forest algorithm.
- Alyoubi, S examined two proposed deep-learning models and achieved the highest accuracy up to 86% by utilizing MARBERT-CNN with the ArCOV19-Rumors dataset.

Table 8. -Details of comparing our model with the baseline models using the AraNews dataset and ArCovid19 Rumors Dataset

Id	Model	Accuracy	F1-Score	Dataset	Study
1	mBert	0.793	0.793	AraNews	Nagoudi, El Moatez Billah [10]
2	XLMR	0.827	0.825		
3	AraBert	**0.872**	0.872		
4	CNN	0.791	0.782		
5	RNN	0.555	0.712	ArCOV19-Rumors	Al-Yahya [11]
6	GRU	0.831	0.838		
7	AraBERT v02	0.945	0.705		
8	AraGPT2	0.813	0.776		
9	QARiB	0.952	0.930		
10	ArBert	**0.958**	0.953		
11	AraBERT v02	0.800	0.886	AraNews	
12	QARiB	**0.800**	0.887		
13	AraGPT2	0.509	0.659		
14	Random Forest	**0.866**	/	AraNews	Aljwari, Fatima [12]
15	Naive Bayes	0.844	/		
16	Logistic Regression	0.859	/		
17	BiLSTM	0.800	0.790	ArCOV19-Rumors	Alyoubi, S[14]
18	MARBERT-CNN	**0.8630**	0.860		
19	**Att-BiLSTM**	**0.905**	0.907	AraNews	**Proposed Model**
		0.969	0.970	ArCOV19-Rumors	

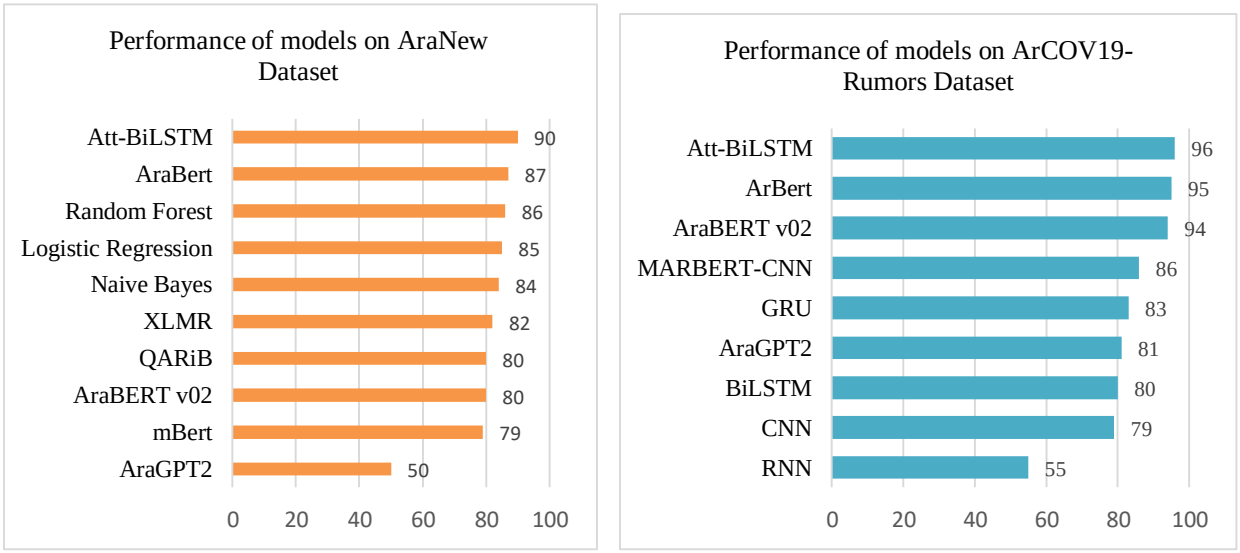


FIGURE 7. - Comparison results of the proposed model and Baseline models on (a) AraNews dataset and (b) ArCOV19-Rumors dataset

From the result of Table (8), firstly, by comparing our model with the baseline models by *Aljwari and Fatima*, our model outperforms all these baselines, achieving an accuracy of 90%. This proves that deep learning-based models perform better than machine learning-based models in the detection of Arabic misinformation. Secondly, we compared our proposed model with baseline models by *"Al-Yahya", "Nagoudi, El Moatez Billah"*, and "*Alyoubi, S"*). With the proposed model achieving accuracy rates of up to 90% and 96% on the AraNews dataset and the ArCOV19-Rumors dataset, respectively, the results clearly demonstrate that our model exhibits superior performance among the techniques investigated in these three studies with regard to the task of Arabic misinformation detection.

5. CONCLUSION

Driven by political or economic motives, technological development and the proliferation of various media outlets contribute to the exacerbation of the spread of misinformation. Recently, the field of misinformation detection has garnered significant interest, with many machine learning methods being proposed for its detection and elimination. In this study, we propose a model based on the attention mechanism for detecting Arabic misinformation. The results of our experiments have led us to several conclusions. In the process of selecting the optimal hyperparameters, the utilization of the Keras Tuner with Bayesian Optimization resulted in the selection of a compatible combination of parameters that lead to an increase in the model's efficiency while also reducing the impact of overfitting on its performance. Moreover, utilizing pre-trained language models like the AraBERT model as textual feature extractors is more effective than other word-embedding methods. Since AraBERT is built based on the transformer architecture, it exploits the attention mechanism performed in the transformer architecture, which efficiently captures the underlying semantic relationships among words in a sentence. We found that the deep learning model on a blend of an attention mechanism and BiLSTM yielded the highest accuracy, up to 0.90, compared to all the existing models across the AraNews dataset. Conversely, our model achieved an accuracy of up to 0.96 on an ArCovid19-Rumors dataset. The attention mechanism's capability to focus the model's attention on relevant parts of the sentence improved the performance of the model in classifying news articles. In future work, we believe that using a combination of multi-head attention with a hybrid model would significantly influence the outcomes. In addition, we intend to predict misinformation based on sentiment analysis.

FUNDING

None

ACKNOWLEDGEMENT

I gratefully acknowledge Mustansiriyah University, Department of Computer Science, for their support and resources in completing my paper.

CONFLICTS OF INTEREST

"The authors have declared that no competing interests exist."

REFERENCES

[1] W. H. Organization, "Infodemics and Misinformation Negatively Affect People's Health Behaviours, New WHO Review Finds. September 1, 2022," ed.

[2] M. Barthel, A. Mitchell, and J. Holcomb, "Many Americans believe fake news is sowing confusion," 2016.

[3] B. Guo, Y. Ding, L. Yao, Y. Liang, and Z. Yu, "The future of misinformation detection: New perspectives and trends," *arXiv preprint arXiv:1909.03654*, 2019.

[4] K. Shu, S. Wang, and H. Liu, "Understanding user profiles on social media for fake news detection," in *2018 IEEE conference on multimedia information processing and retrieval (MIPR)*, 2018: IEEE, pp. 430–435.

[5] M. Cantarella, N. Fraccaroli, and R. Volpe, "Does fake news affect voting behaviour?," *Research Policy*, vol. 52, no. 1, p. 104628, 2023.

[6] P. N. Petratos, "Misinformation, disinformation, and fake news: Cyber risks to business," *Business Horizons*, vol. 64, no. 6, pp. 763–774, 2021.

[7] A. Zareie and R. Sakellariou, "Minimizing the spread of misinformation in online social networks: A survey," *Journal of Network and Computer Applications*, vol. 186, p. 103094, 2021.

[8] S. F. Sabbeh and S. Y. Baatwah, "ARABIC NEWS CREDIBILITY ON TWITTER: AN ENHANCED MODEL USING HYBRID FEATURES," *journal of theoretical & applied information technology*, vol. 96, no. 8, 2018.

[9] H. Saadany, E. Mohamed, and C. Orasan, "Fake or real? A study of Arabic satirical fake news," *arXiv preprint arXiv:2011.00452*, 2020.

[10] E. M. B. Nagoudi, A. Elmadany, M. Abdul-Mageed, T. Alhindi, and H. Cavusoglu, "Machine generation and detection of Arabic manipulated and fake news," *arXiv preprint arXiv:2011.03092*, 2020.

[11] M. Al-Yahya, H. Al-Khalifa, H. Al-Baity, D. AlSaeed, and A. Essam, "Arabic fake news detection: comparative study of neural networks and transformer-based approaches," *Complexity*, vol. 2021, pp. 1–10, 2021.

[12] F. Aljwari, W. Alkaberi, A. Alshutayri, E. Aldhahri, N. Aljojo, and O. Abouola, "Multi-scale Machine Learning Prediction of the Spread of Arabic Online Fake News," *Postmodern Openings*, vol. 13, no. 1 Sup1, pp. 01–14, 2022.

[13] K. M. Fouad, S. F. Sabbeh, and W. Medhat, "Arabic Fake News Detection Using Deep Learning," *Computers, Materials & Continua*, vol. 71, no. 2, 2022.

[14] S. Alyoubi, M. Kalkatawi, and F. Abukhodair, "The Detection of Fake News in Arabic Tweets Using Deep Learning," *Applied Sciences*, vol. 13, no. 14, p. 8209, 2023.

[15] M. F. Ibrahim, M. A. Alhakeem, and N. A. Fadhil, "Evaluation of Naïve Bayes classification in Arabic short text classification," *Al-Mustansiriyah J. Sci.*, vol. 32, no. 4, pp. 42–50, 2021.

[16] Z. A.-W. Salman, "Text Summarizing and Clustering Using Data Mining Technique," *Al-Mustansiriyah Journal of Science*, vol. 34, no. 1, pp. 58–64, 2023.

[17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[18] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[19] W. Antoun, F. Baly, and H. Hajj, "Arabert: Transformer-based model for arabic language understanding," *arXiv preprint arXiv:2003.00104*, 2020.

[20] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE access*, vol. 6, pp. 1662–1669, 2017.

[21] F. Shahid, A. Zameer, and M. Muneeb, "Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM," *Chaos, Solitons & Fractals*, vol. 140, p. 110212, 2020.

[22] N. Rai, D. Kumar, N. Kaushik, C. Raj, and A. Ali, "Fake News Classification using transformer based enhanced LSTM and BERT," *International Journal of Cognitive Computing in Engineering*, vol. 3, pp. 98–105, 2022.

[23] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[24] M. H. Al-Tai, B. M. Nema, and A. Al-Sherbaz, "Deep Learning for Fake News Detection: Literature Review," *Al-Mustansiriyah Journal of Science*, vol. 34, no. 2, pp. 70–81, 2023.

[25] M. Fazil, A. K. Sah, and M. Abulaish, "Deepsbd: a deep neural network model with attention mechanism for socialbot detection," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 4211–4223, 2021.

[26] F. Haouari, M. Hasanain, R. Suwaileh, and T. Elsayed, "ArCOV19-rumors: Arabic COVID-19 twitter dataset for misinformation detection," *arXiv preprint arXiv:2010.08768*, 2020.

[27] B. M. Nema and S. J. Mohammed, "Secure Location Privacy Transmitting Information on Cellular Networks", Iraqi Journal of Science, vol. 63, no. 11, pp. 5004–5014, Nov. 2022.