



ISSN: 0067-2904

Attention Mechanism Based on a Pre-trained Model for Improving Arabic Fake News Predictions

Tahseen A. Wotaifi^{1*}, Ban N. Dhannoon²

¹College of Information Technology, University of Babylon, Hillah, Babil, Iraq

²Department of Computer Science, College of Science, Al-Nahrain University, Baghdad, Iraq,

Received: 19/9/2022

Accepted: 4/2/2023

Published: 30\11\2023

Abstract

Social media and news agencies are major sources for tracking news and events. With these sources' massive amounts of data, it is easy to spread false or misleading information. Given the great dangers of fake news to societies, previous studies have given great attention to detecting it and limiting its impact. As such, this work aims to use modern deep learning techniques to detect Arabic fake news. In the proposed system, the attention model is adapted with bidirectional long-short-term memory (Bi-LSTM) to identify the most informative words in the sentence. Then, a multi-layer perceptron (MLP) is applied to classify news articles as fake or real. The experiments are conducted on a newly launched Arabic dataset called the Arabic Fake News Dataset (AFND). The AFND dataset contains exactly 606912 news articles collected from multiple sources, so it is suitable for deep learning requirements. Both simple recurrent neural networks (S-RNN), long short-term memory (LSTM), and gated recurrent units (GRU) are used for comparison. According to evaluation criteria, our proposed model achieved an accuracy of (0.8127), which is the best and highest accuracy among the deep learning methods used in this work. Moreover, the performance of our proposed model is better compared to previous studies, which used the AFND.

Keywords: Arabic Fake News, Attention Model, Bi-LSTM, GRU, and Deep Learning.

آلة الانتباه بالاعتماد على النماذج المدربة مسبقاً لتحسين تنبؤات الاخبار العربية المزيفة

تحسين علي حسن^{1*} ، بان نديم ذنون²

¹كلية تكنولوجيا المعلومات، جامعة بابل، حلة. بابل. عراق

²قسم علوم الحاسبات، كلية العلوم، جامعة النهرين، بغداد، عراق

الخلاصة

تعد وسائل التواصل الاجتماعي ووكالات الأنباء من المصادر الرئيسية لتتبع الأخبار والأحداث. مع البيانات الضخمة لهذه المصادر ، من السهل نشر أخبار غير موثوقة أو أخبار مزيفة. نظراً للمخاطر الكبيرة للأخبار الكاذبة على المجتمعات ، فقد أولت الدراسات السابقة اهتماماً كبيراً لكشفها والحد من تأثيرها. على هذا النحو ، يهدف هذا العمل إلى استعمال تقنيات التعلم العميق الحديثة للكشف عن الأخبار العربية المزيفة. في النظام المقترح ، تم تكيف نموذج الانتباه مع الذاكرة ثنائية الاتجاه طويلة المدى (Bi-LSTM) لتحديد الكلمات

*Email: tahseen.ali@uobabylon.edu.iq

الأكثر إفادة في الجملة. بعد ذلك ، يتم تطبيق منظور متعدد الطبقات (MLP) لتصنيف المقالات الإخبارية على أنها مزيفة أو حقيقية. أجريت التجارب على مجموعة بيانات عربية تم إطلاقها حديثاً تسمى مجموعة بيانات الأخبار الكاذبة العربية "AFND". تحتوي مجموعة بيانات AFND بالضبط على (606912) مقالة إخبارية تم جمعها من مصادر متعددة ، لذا فهي مناسبة لمتطلبات التعلم العميق. كلاً من الشبكة العصبية المتكررة البسيطة (S-RNN) يتم استخدام RNN ، والذاكرة طويلة المدى (LSTM) ، ووحدة البوابات المتكررة (GRU) تم استعمالها للمقارنة. وفقاً لمعايير التقييم ، فقد حقق نموذجنا المقترح دقة قدرها (0.8127) ، وهي أفضل وأعلى دقة بين طرق التعليم العمق المستعملة في هذا العمل. علاوة على ذلك ، فإن أداء نموذجنا المقترح أفضل مقارنة بالدراسات السابقة التي استعملت AFND..

1. Introduction

"Fake news" or "uncredible news" is defined as information or a claim that is verified to be incorrect. This news is usually posted and circulated on the internet and social media platforms for the purpose of misleading people and changing facts [1]. As a result of the lack of tight monitoring of these platforms, the spread of fake news is easy and then has wide resonance as publishers or users choose sensitive topics such as politics, religion, and economics [2]. As a result, the implications of this news are dire, and it may endanger countries and community peace [3].

Predicting the credibility of a news article, story, or tweet and checking whether it is true or false is called "fake news detection" (FND) [4]. Other terms close in concept to FND are rumor detection, disinformation classification, stance classification, and misleading information detection [5]. In the last few years, researchers and the scientific community have paid great attention to FND. Many studies have been conducted on identifying fake news in English; however, detection of Arabic fake news is still in its infancy [6]. The reason for this is that there is no comprehensive Arabic dataset that is approved by researchers. Recently, a huge dataset was launched, contributed by a number of researchers, called the AFND. This dataset contains a huge collection of news articles collected from many sources, and each article is classified as not credible, credible, or undecided [7].

Manually determining whether a news article is fake or real is slow, tedious, and impractical [8]. The recent trends in the research community are to include machine learning (ML) and deep learning (DL) and then build models to predict whether the news is true or false [9]. Moreover, this work aims to use the attention model to select relevant features. Then, Bi-LSTM is used to define context by measuring the long and short sequences. Finally, MLP is applied to predict Arab fake news. The results showed that the performance of our proposed model is better compared to previous studies that used the AFND. Briefly, the main contributions of this study can be summarized as follows:

1. Creating an embedding matrix for AFND helps the model train faster by avoiding searching in an embedding matrix containing millions of words (the embedding matrix of pre-trained models). This matrix also aims to improve accuracy by finding embedding vectors for unrepresented words.
2. Building a hybrid model that uses the Bi-LSTM to process long sequences (long news articles). Then the attention model is applied to the output of the bi-LSTM to determine the most informative words. Finally, classify articles by using MLP.

2. Related Work

Due to the problems resulting from fake news, many studies have been conducted to limit its spread. Some studies focus on the textual content and context for detecting fake news,

whereas other studies concern themselves with the features of the user or writer. Despite researchers' clear contributions in presenting works to solve the problem of untrustworthy news, few of them were interested in detecting Arabic fake news. There are few such studies because there is no public dataset in Arabic for researchers to use. The researchers' trends were on the Twitter platform in collecting data; however, a large dataset recently appeared in Arabic called AFND. This work uses this dataset for building prediction models, and it was found that our proposed methodology is better in terms of performance than the previous study conducted on the same dataset.

According to AFND, the authors in [1] have presented a study to predict Arabic fake news. In the detection task, this study uses ML and DL. The study shows that DL techniques are better in terms of accuracy and performance in predicting fake news. Moreover, multi-classifier and binary-classifier tests were performed, and in both approaches, the capsule network model achieved the best accuracy (0.709 accuracy for multi-classifier and 0.79 accuracy for binary classifier). Our proposed model is built using the AFND, and it was found through evaluation metrics that the performance of the proposed model is better compared to this study.

As for the study in [2], the Twitter platform was adopted for data collection by the authors. In this study, the authors focused on user features, and the most important ones were identified through a fuzzy model. On the other hand, a TF-IDF is used to represent words as features. In the prediction stage, a modified machine learning model was proposed, and its results were better compared to other models (the accuracy of this model is 0.89).

In [3], the authors use the AraNews dataset to build the models. This dataset contains a large collection of news articles covering approximately 15 Arab countries. The prediction of fake news was performed using traditional ML techniques, and the random forest model achieved an accuracy equal to 866, which is better compared to the models used in this work.

3. Theoretical Background

In the theoretical background section, a brief overview of deep learning techniques is presented. Specifically, the dataset is explained in Section 3.1, and then RNN, LSTM, and GRU networks are described in Sections 3.2, 3.3, and 3.4, respectively. After that, the attention mechanism and transfer learning are presented in Sections 3.5 and 3.6, respectively.

3.1 Dataset

AFND is a newly launched huge Arabic dataset that is suitable for traditional ML techniques and DL techniques. This dataset consists of 606912 records collected from 134 news websites in 19 Arabic countries. Using the Arabic fact-check platform (Misbar), each news article is classified as not credible, credible, or undecided. Table 1 views the statistical information for AFND, and Figure 1 explains the percentages of the news articles for each class [1].

Table 1: statistical information of AFND

Statistics	Undecided	Credible	Not Credible
No. of Articles	232396	207310	167233
No. of Sources	31	52	51

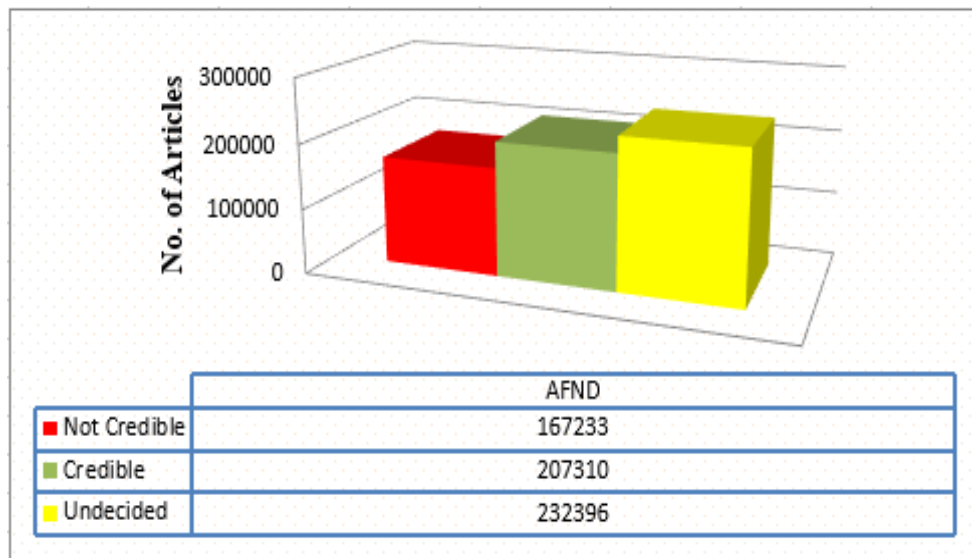


Figure 1: Number of Articles for Each Class

3.2 Recurrent Neural Network

RNN is a feed-forward neural network that handles variable-length text input via a recurrent hidden layer [10]. The main idea behind this network is that the input of the current state depends on the output of the previous state, and thus the context is preserved. Usually, the RNN is not suitable for long news articles due to the gradient vanishing problem [10].

Figure 2 illustrates the structure of RNN. x_t represents the input vector at time t . h_t is the output of the network node (hidden state) at time t . The hidden state is computed based on the previous hidden state and the input vector at the same time [11].

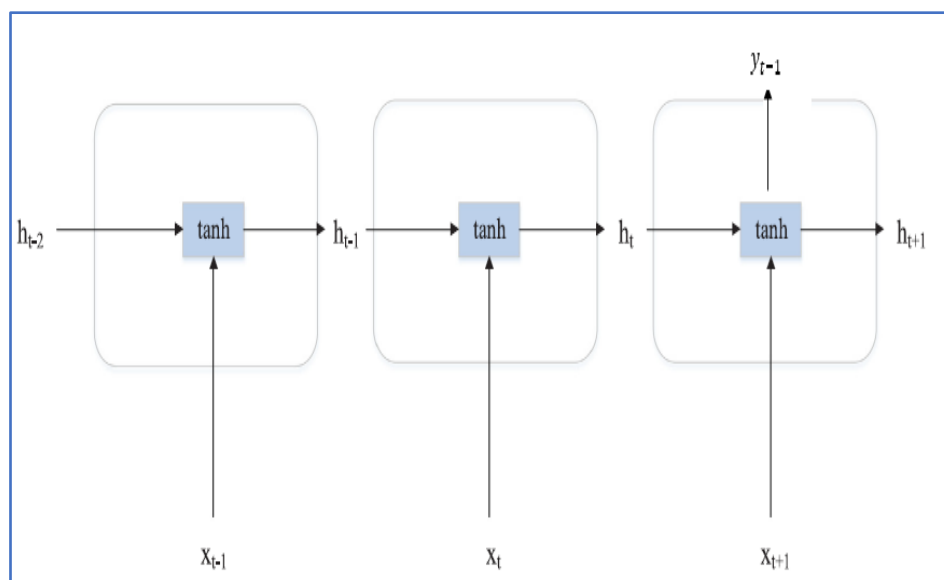


Figure 2: RNN Framework

The outputs of an RNN are obtained based on f_H and f_O which represent the activation functions of the hidden state and output state, respectively, as shown in Equations 1 and 2 [12].

$$h_t = f_H(W_{IH}x_t + W_{HH}h_{t-1} + b_h) \tag{1}$$

$$y_t = f_O(W_{OH} + b_y) \tag{2}$$

Where: W_{OH} , W_{HH} , and W_{IH} are the weight vectors corresponding to output, hidden, and input, respectively.

3.3 Long-Short Term Memory

The LSTM network shown in Figure 3 is a special case of an RNN that aims to learn long-term dependencies [13]. This network is a suitable solution for handling the problem of vanishing gradient that the RNN suffers from [14]. LSTM uses an input gate, forget gate, and output gate to regulate the flow of information. These three gates determine which text sequences are significant enough to keep or throw away, and thus it is easy to predict long sequences [15].

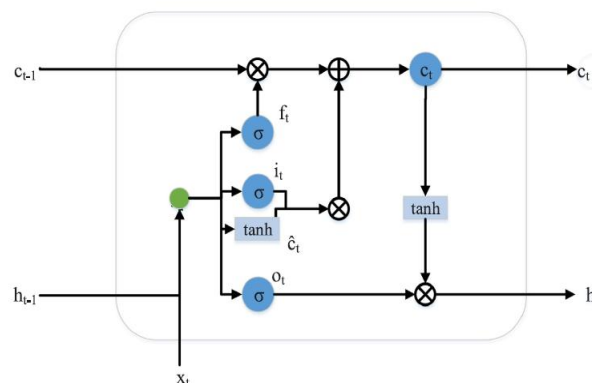


Figure 3: LSTM Framework

The output of hidden state " h_t ", cell state " c_t " and output state o_t for the LSTM neuron is computed as follows [16]:

$$FG_t = \sigma(W_{FG} \cdot [h_{t-1}, x_t] + b_{FG}) \quad (3)$$

$$IG_t = \sigma(W_{IG} \cdot [h_{t-1}, x_t] + b_{IG}) \quad (4)$$

$$OG_t = \sigma(W_{OG} \cdot [h_{t-1}, x_t] + b_{OG}) \quad (5)$$

$$\check{c}_t = \tanh(W_{CG} \cdot [h_{t-1}, x_t] + b_c) \quad (6)$$

$$c_t = FG_t \otimes c_{t-1} + \check{c}_t \quad (7)$$

$$h_t = OG_t \otimes \tanh(c_t) \quad (8)$$

Where: FG_t is forget gate at timestep t . IG_t and OG_t are input and output gates at timestep t . w , h , x , and c are vectors. \tanh and σ is activation functions.

3.4 Gate Recurrent Unit

LSTM was proposed for processing text and has excellent capabilities for memorizing long input sequences. However, the complex structure of this network makes it take a long time. This problem was addressed by proposing the GRU network, which has a simpler structure [17].

In the GRU structure shown in Figure 4, the cell state c_t and hidden state h_t are merged into one. Thus, there are two gates: the Reset Gate and The Update Gate. When the value of the Update Gate is large, the status information of the previous time increases. On the other hand, the Reset Gate controls the amount of information from the previous state [18].

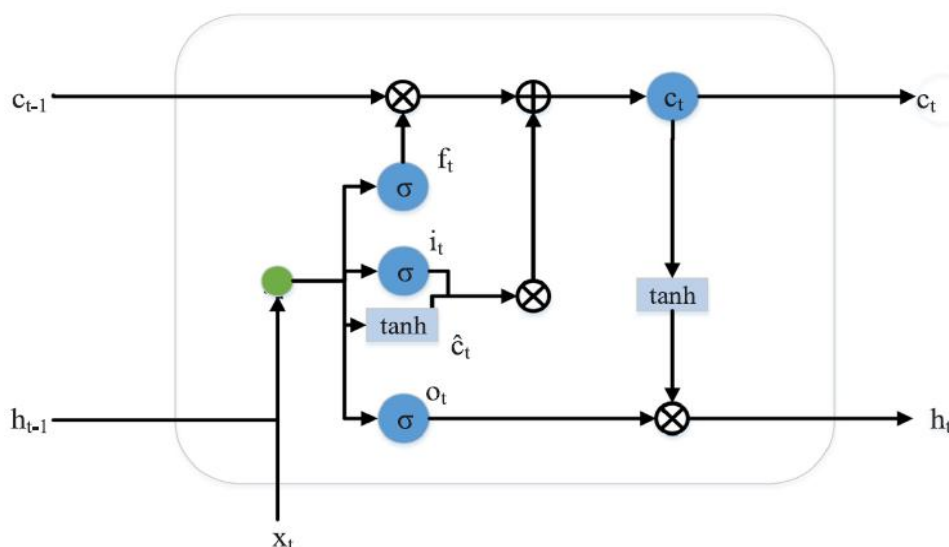


Figure 4: GRU Framework

In the architecture of the GRU node illustrated in Figure 5, update equations are computed as follows [19]:

$$UG_t = \sigma(W_{xUG}x_t + W_{hUG}h_{t-1} + b_{UG}) \quad (9)$$

$$RG_t = \sigma(W_{xRG}x_t + W_{hRG}h_{t-1} + b_{RG}) \quad (10)$$

$$\check{c}_t = \tanh(W_{xc}x_t + W_{hc}(UG_t \otimes h_{t-1}) + b_c) \quad (11)$$

$$c_t = (1 - RG_t) \otimes c_{t-1} + RG_t \otimes \check{c}_t \quad (12)$$

$$h_t = c_t \quad (13)$$

Where: UG_t is the update gate at timestep t . RG_t is reset gates at timestep t . w , h , x , and c are vectors. \tanh and σ is activation functions.

3.5 Attention Model

The main idea of the attention model is to let the decoder focus on relevant words in the input sequence by assigning a weight to each word [20]. In other words, this model gives more attention to certain words in the text and ignores other words [21]. The attention model is divided into the step-by-step computations of the attention scores, the attention weights, and the context vector [22].

1- Attention Scores: As in Equation 14, the attention score (e) is computed from two inputs: the encoded hidden states (h_t) and the previous decoder output (s_{t-1}). The score value indicates how well the input sequence matches the current output at position t . The attention model is represented through a function (f). This function is implemented by a feedforward network [23].

$$e_{t,i} = f(s_{t-1}, h_t) \quad (14)$$

2- Attention Weights: Attention weights are computed based on attention score and SoftMax operation. Equation 15 shows the calculation of these weights [23].

$$\alpha_{t,i} = \text{softmax}(e_{t,i}) \quad (15)$$

3- Context Vector: At each time step, the context vector is fed into the decoder. Then, a unique context vector is computed by a weighted sum of all encoder hidden states (T) as follows [23].

$$c_t = \sum_{i=1}^T \alpha_{t,i} \cdot h_t \quad (16)$$

3.6 Transfer Learning

Transfer learning is a set of pre-trained models intended to represent words. These models were trained on millions of words in order to represent each word as a feature or vector that indicates the semantic meaning of that word. The two most important pre-trained models that were trained on an Arabic dataset are GloVe and Fast Text [24].

GloVe is unsupervised learning developed by Stanford University for word representation. In order to obtain vector representations of words, the words are assigned to a meaningful space, and then the semantic similarity is used to represent each word [25]. According to this model, the embedding dimension (embedding value) for each word is 300. Similarly, the Fast Text pre-trained model, which was developed by Facebook's AI Research Lab, assigns a vector or feature to each word of size equal to 300 embedding values. This module is an open-source library that uses the concepts of deep learning and natural language processing for learning word representation and efficient text classification [26].

4. Methodology

This section describes the mechanism and techniques used to construct the proposed model. Given the significant problems and threats posed by fake news, the methodology proposed in this work aims to mitigate the negative impact of these issues by detecting untrustworthy news (fake news). As shown in Figure 5, the proposed solution to detect this news is achieved through five stages: 1) All news articles in the AFND are pre-processed; 2) words or tokens are represented as features (vectors) through pre-trained models; 3) BiLSTM is used to make the model capable of capturing short or long sequences; 4) the more informative words are determined through the attention model; and 5) finally, the artificial neural network is used to classify the news article as credible or not credible.

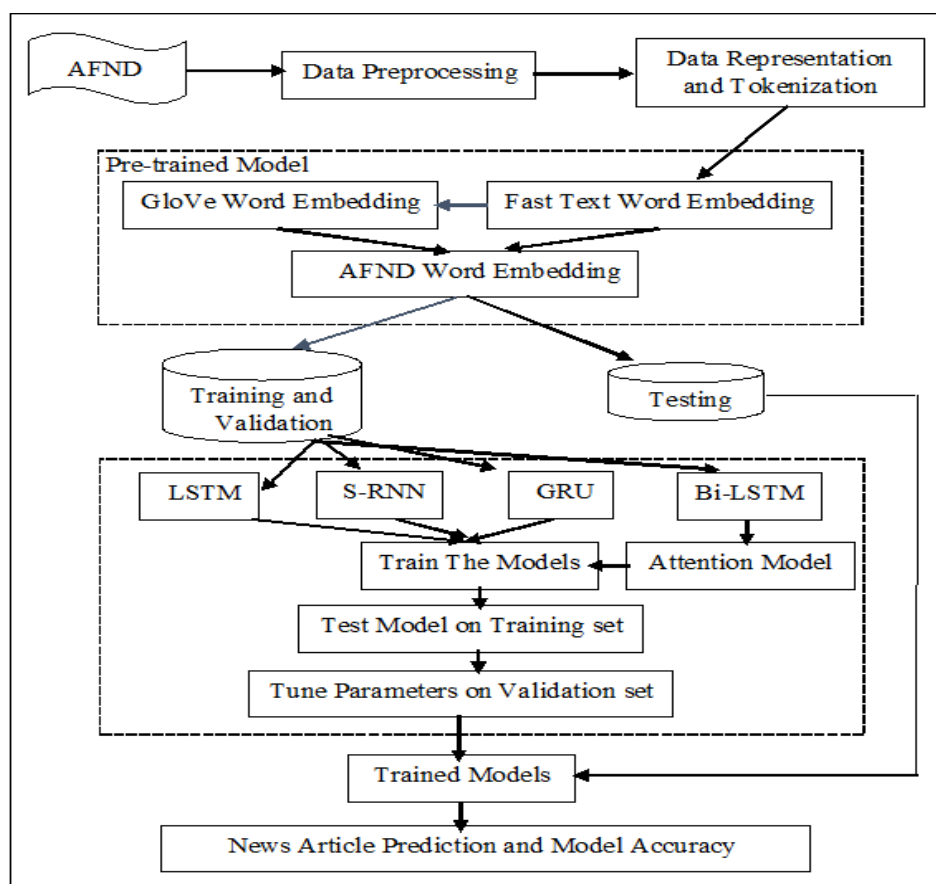


Figure 5: Proposed Methodology

In this work, the binary classification has been considered by focusing on credible and not credible articles in the AFND, whereas undecided articles have been ignored. The rest of the dataset is 103 JSON files converted into one CSV file. All news articles in this file are pre-processed. Preprocessing has included removing punctuation, non-Arabic words, and stop words. Then the Arabic stemming is performed to reduce the dimensions of the words. The news article and its title are merged and then divided into the list of tokens. The last step of this stage is to perform the padding and truncating technique in order to have an equal size for all the news articles.

FastText and Global Vector (GloVe) pre-trained models are used for word representation. In our proposed model, the AFND embedding matrix is built based on these models that represent the initial weights of the embedding layer. This matrix contains the embedding values for the words in the AFND, where each vector of this matrix represents a word in the AFND. The fast text is mainly used for word representation. In contrast, the GloVe is used to represent the words in the AFND that do not have vectors in the Fast Text. In addition, any word that was not represented in both models is assigned a vector based on the word closest to it. The purpose of this matrix is to improve the accuracy of the model by solving the out-of-vocabulary problem. Also, execution time is improved by searching a matrix containing a few vectors instead of the pre-trained model matrix containing millions of vectors.

Then, the output of the embedding layer is fed to the bi-LSTM layer. As shown in Figure 6, Bi-LSTM is used to capture the context and process the long and short sequences in both directions (forward and backward). When using the LSTM network to predict fake news, the traditional method is to use an output layer (the classification layer) with the sigmoid function

to classify news articles. Instead, the output of this network is fed into the attention layer. The reason behind this procedure is that the LSTM network gives all the words in the news article equal importance. In other words, the problem with bi-LSTM is that it cannot identify the more informative words. To address this problem, the attention model was applied to the outputs of BiLSTM to make the system capable of paying less or more attention to words in the AFND. The more informative words are obtained through Equations 14, 15, and 16.

Finally, MLP with five layers has been applied to predict fake news articles. Two of these layers are fully connected. After each of these two layers, a dropout layer was added to avoid the problem of overfitting. Finally, the output layer (a classification layer) is applied with a sigmoid activation function to classify the news articles. Figure 6 shows the main structure of the proposed model.

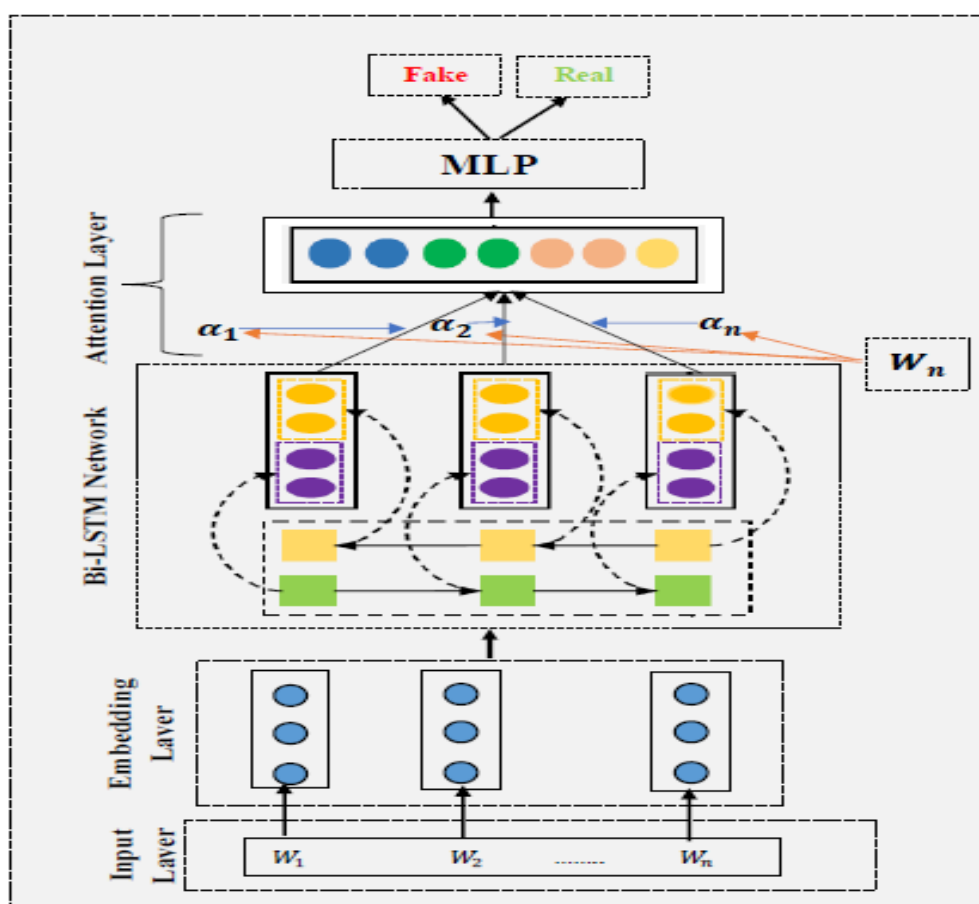


Figure 6: The Structure of Proposed Model

5. Results and Discussion

This section discusses the performance results of our proposed model in Arabic fake news prediction by conducting experiments on the AFND. At first, all codes are written in Python using Keras, and all experiments are performed on a Core i7 processor and 8GB of RAM. The AFND is used to measure the quality of the proposed methodology. The dataset is available online. Each news article in AFND consists of a title, text, publication date, and target class as undecided, not credible, or credible. In this work, news articles classified as undecided are ignored, and therefore the proposed model is a binary classification.

In order to train the models, the title and text of each news article in AFND are combined. All news articles in this dataset are pre-processed by removing symbols, non-Arabic words,

and noise. To avoid overfitting and underfitting, the AFND is partitioned into 60% for the training set, 20% for the validation set, and 20% for the testing set. The idea behind the validation set is to tune the hyperparameters of the model. Table 1 provides more details about this partition.

Table 1: Dataset Details.

Class Label	Total	Training	Testing	Validation
Credible (Not Fake)	207320	124392	41464	41464
Not Credible (Fake)	167233	100341	33446	33446

In the proposed system, the model is implemented by Keras. Nine layers are built: The first layer is an untrainable embedding layer based on pre-trained models. Two layers of Bi-LSTM are added to measure long-term dependencies at each layer (64 and 32 units, respectively). To identify more informative words (relevant features), the attention layer is used. Then five layers were added to classify the news article as fake or real: Two fully connected layers (with 64 and 32 units per layer, respectively), two dropout layers to avoid overfitting, and a classification layer (output layer). The sigmoid activation function is used in the output layer, while the Rectified Linear Unit Activation Function (ReLU) has been used with all the remaining layers.

The important aspect of the proposed methodology is the setting and tuning of hyperparameters. The Fast Text and GloVe pre-trained models are used for word representation, so the embedding dimension for each token is 300. Based on the average number of words in each news article for the AFND, the maximum length is determined to be 55. The probability of dropping neurons at the two dropout layers is determined to be 0.20 and 0.25, respectively, whereas the L2 regularization parameter is 0.01. The learning rate is set to be 0.0005 based on trial and error. Finally, the batch size is set to 128 and 15 training epochs are identified. The main parameters are reviewed in Table 2.

Table 2: The Main Parameters

Name	Size
Bach Size	128
Epoch	15
Max-Length	55
Embedding Dim.	300
L2 Reg.	0.01
Dropout	0.2 and 0.25
LR	0.0005

In the model training phase, all parameters are updated using the Adam optimizer. Then, evaluation metrics are used to determine the quality and performance of these models. As shown in Equation (17), prediction accuracy, which is calculated as a ratio of the number of correctly predicted articles to the total number of articles in the testing set, was used as an evaluation metric [18]. Based on their accuracy, the proposed models are compared with previous works. Few studies have been conducted on the AFND; however, when compared to the work in [1], which used the same dataset, our proposed model outperforms it in terms of prediction accuracy. Table 4 presents the results of the proposed models and their comparison with previous work.

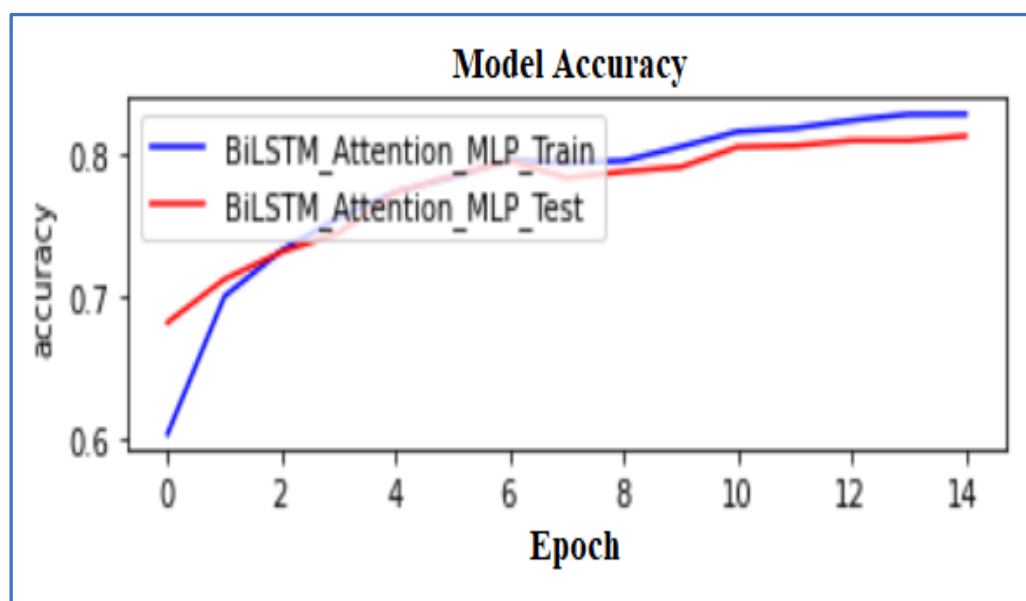
$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + Fals\ Positive + False\ Negative} \quad (17)$$

Table 4: Summary of Results and Comparisons

Model	Training Acc.	Validation Acc.	Testing Acc.	Time
Attention_Bi-LSTM	0.8285	0.8127	0.8173	39 sec.
LSTM	0.7981	0.7805	0.7812	
GRU	0.7936	0.7756	0.7801	35 sec.
RNN	0.7803	0.7512	0.7695	31 sec.
Study in [1]	-----	-----	0.79	-----

Few studies were conducted on the AFND because this dataset was recently launched. As shown in Table 4, the authors in Study [1] used the same dataset (AFND) as the one used in this study, and the best accuracy achieved was 0.79. Our proposed model that uses the attention mechanism with the Bi-LSTM is superior to this study, which was used for comparison. The proposed model outperforms in terms of accuracy also to the baseline models: Simple RNN, LSTM, and GRU (see Table 4).

To display the results of the proposed model visually, Figure 7 shows the convergence of the training accuracy and testing accuracy, Figure 8 explains the model loss for training and testing, and Figure 9 illustrates the confusion matrix.

**Figure 7:** Convergence of Model Accuracy

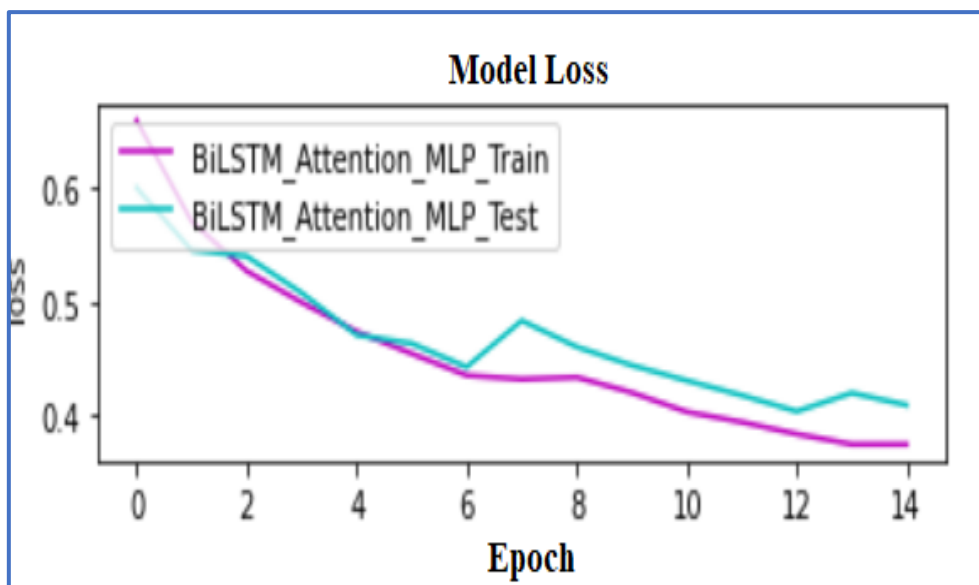


Figure 8: The Decrease in Model Loss

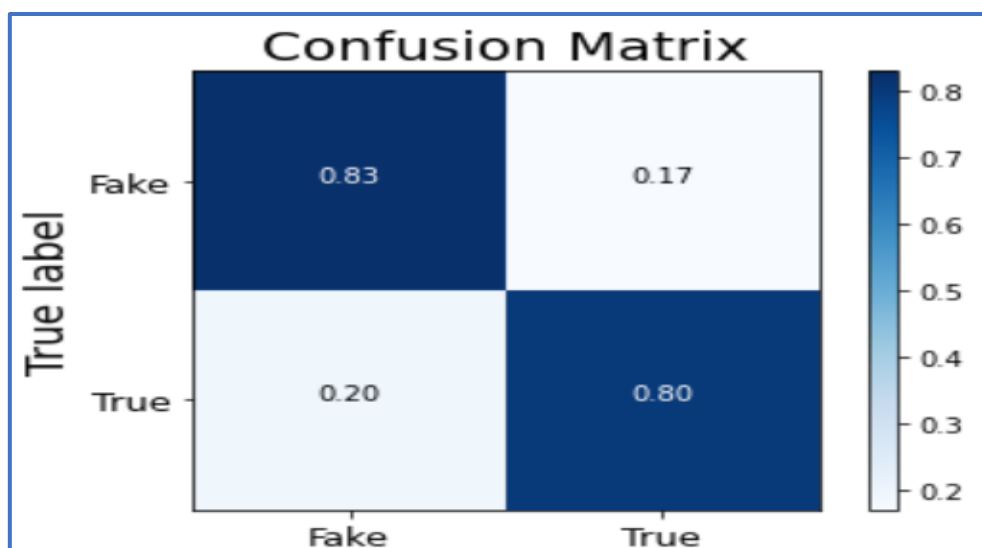


Figure 9: Confusion Matrixes of the Proposed Model

6. Conclusion

In the last few years, it has been observed that the amount of data circulated via the Internet and social media is increasing dramatically. Thus, transferring fake news through such data and misleading people is easy for the user or publisher. On the other hand, the research and academic communities have shown significant interest in mitigating the consequences of spreading fake news through the use of ML and DL. Many studies have been conducted to detect English fake news; however, the interest in detecting Arabic fake news is undeveloped. This work aims to predict Arabic fake news using modern DL techniques.

In our proposed model, both Fast Text and GloVe pre-trained models are used to construct a matrix of vectors for AFND. The goal behind this matrix is to speed up model execution and improve accuracy. The bi-LSTM layer is applied to the output of the embedding layer. This makes the model capable of addressing arbitrary and short sequences. Then, the more informative words are identified by adding an attention layer. The output of this layer is a single vector representing the most important words in the sentence. Finally, the news articles

are classified as fake or real using MLP. Based on the accuracy metric, it was found that the proposed model achieves the best performance in detecting fake news compared with previous work conducted on AFND.

In future works, bidirectional GRU can be used with the Attention Mechanism to process long and short sentences. In our plans, we intend to improve prediction accuracy by combining a convolutional neural network with an attention model to identify context and relevant words.

References

- [1] A. Khalil, M. Jarrah, M. Aldwairi, and Y. Jararweh, "Detecting arabic fake news using machine learning," in *2021 Second International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, pp. 171–177, 2021.
- [2] T. A. Wotaifi and B. N. Dhannoon, "Improving Prediction of Arabic Fake News Using Fuzzy Logic and Modified Random Forest Model," *Karbala Int. J. Mod. Sci.*, vol. 8, no. 3, pp. 477–485, 2022.
- [3] F. Aljwari, W. Alkaber, A. Alshutayri, E. Aldahri, N. Aljojo, and O. Abouola, "Multi-scale Machine Learning Prediction of the Spread of Arabic Online Fake News," *Postmod. Openings*, vol. 13, no. 1 Sup1, pp. 1–14, 2022.
- [4] 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, 2021.
- [5] P. Bahad, P. Saxena, and R. Kamal, "Fake news detection using bi-directional LSTM-recurrent neural network," *Procedia Comput. Sci.*, vol. 165, pp. 74–82, 2019.
- [6] W. Shishah, "JointBert for Detecting Arabic Fake News," *IEEE Access*, vol. 10, pp. 71951–71960, 2022.
- [7] T. A. Wotaifi and E. S. Al-Shamery, "Modified Random Forest based Graduates Earning of Higher Education Mining," *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.*, vol. 12, pp. 56–65, 2020.
- [8] T. A. Wotaifi and E. S. Al-Shamery, "Fuzzy-Filter Feature Selection For Envisioning The Earnings Of Higher Education Graduates," *Compusoft*, vol. 7, no. 12, pp. 2969–2975, 2018.
- [9] M. Al-Fawa'rah, Z. Ashi, and M. T. Jafar, "Detecting Malicious DNS Queries Over Encrypted Tunnels Using Statistical Analysis and Bi-Directional Recurrent Neural Networks," *Karbala Int. J. Mod. Sci.*, vol. 7, no. 4, pp. 4, 2021.
- [10] M. E. Basiri, S. Nemati, M. Abdar, E. Cambria, and U. R. Acharya, "ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis," *Futur. Gener. Comput. Syst.*, vol. 115, pp. 279–294, 2021.
- [11] Z. Hussien and B. N. Dhannoon, "Anomaly Detection Approach Based on Deep Neural Network and Dropout," *Baghdad Sci. J.*, vol. 17, no. 2 (SI), pp. 701-709, 2020.
- [12] N. A. Hamzah and B. N. Dhannoon, "The Detection of Sexual Harassment and Chat Predators Using Artificial Neural Network," *Karbala Int. J. Mod. Sci.*, vol. 7, no. 4, pp. 6, 2021.
- [13] S. Gao *et al.*, "Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation," *J. Hydrol.*, vol. 589, pp. 1-11, 2020.
- [14] F. A. Abdulghani and N. A. Z. Abdullah, "A survey on Arabic text classification using deep and machine learning algorithms," *Iraqi J. Sci.*, vol. 63, no. 1, pp. 409–419, 2022.
- [15] M. A. H. Ashour, "Optimized Artificial Neural network models to time series," *Baghdad Sci. J.*, pp. 899, 2022.
- [16] G. Gravanis, A. Vakali, K. Diamantaras, and P. Karadais, "Behind the cues: A benchmarking study for fake news detection," *Expert Syst. Appl.*, vol. 128, pp. 201–213, 2019.
- [17] J. Du, C.-M. Vong, and C. L. P. Chen, "Novel efficient RNN and LSTM-like architectures: Recurrent and gated broad learning systems and their applications for text classification," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1586–1597, 2020.
- [18] T. A. Wotaifi and B. N. Dhannoon, "An Effective Hybrid Deep Neural Network for Arabic Fake News Detection," *Baghdad Science Journal*, Jan. 2023, doi: 10.21123/bsj.2023.7427.
- [19] N. Albadi, M. Kurdi, and S. Mishra, "Investigating the effect of combining GRU neural networks

- with handcrafted features for religious hatred detection on Arabic Twitter space,” *Soc. Netw. Anal. Min.*, vol. 9, no. 1, pp. 1–19, 2019.
- [20] J. Li, K. Jin, D. Zhou, N. Kubota, and Z. Ju, “Attention mechanism-based CNN for facial expression recognition,” *Neurocomputing*, vol. 411, pp. 340–350, 2020.
- [21] H. Fukui, T. Hirakawa, T. Yamashita, and H. Fujiyoshi, “Attention branch network: Learning of attention mechanism for visual explanation,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10705–10714, 2019.
- [22] H. A. Ahmed and E. A. Mohammed, “Detection and Classification of The Osteoarthritis in Knee Joint Using Transfer Learning with Convolutional Neural Networks (CNNs),” *Iraqi Journal of Science*, pp. 5058–5071, Nov. 2022.
- [23] H. Lv, J. Chen, T. Pan, T. Zhang, Y. Feng, and S. Liu, “Attention mechanism in intelligent fault diagnosis of machinery: A review of technique and application,” *Measurement*, pp. 1-18, 2022.
- [24] Md. A. H. Wadud, M. F. Mridha, and M. M. Rahman, “Word Embedding Methods for Word Representation in Deep Learning for Natural Language Processing,” *Iraqi Journal of Science*, pp. 1349–1361, Mar. 2022.
- [25] C. Xiao, X. Hu, Z. Liu, C. Tu, and M. Sun, “Lawformer: A pre-trained language model for chinese legal long documents,” *AI Open*, vol. 2, pp. 79–84, 2021.
- [26] S. Yang and R. Tang, “Learning Knowledge Uncertainty from the Pretrained Language Model,” *2021 the 6th International Conference on Information Systems Engineering*, Nov. 2021.