A Fake News Detection System based on Combination of Word Embedded Techniques and Hybrid Deep Learning Model

Mohamed-Amine OUASSIL[1], Bouchaib CHERRADI[2], Soufiane HAMIDA[3], Mouaad ERRAMI[4], Oussama EL GANNOUR[5], Abdelhadi RAIHANI[6]

Electrical Engineering and Intelligent Systems (EEIS) Laboratory, ENSET of Mohammedia
Hassan II University of Casablanca, Mohammedia 28830, Morocco[1,2,3,4,5,6]
STIE Team, CRMEF Casablanca-Settat, Provincial Section of El Jadida, El Jadida 24000, Morocco[2]

***Abstract*—At present, most people prefer using different online sources for reading news. These sources can easily spread fake news for several malicious reasons. Detecting this unreliable news is an important task in the Natural Language Processing (NLP) field. Many governments and technology companies are engaged in this research field to prevent the manipulation of public opinion and spare people and society the huge damage that can result from the spreading of misleading information on online social media. In this paper, we present a new deep learning method to detect fake news based on a combination of different word embedding techniques and a hybrid Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BILSTM) model. We trained the classification model on the unbiased dataset WELFake. The best method was a combination of a pre-trained Word2Vec CBOW model and a Word2Vec Skip-Word model with a CNN on BILSTM layers, yielding an accuracy of up to 97%.**

***Keywords*—Deep learning (DL); Bidirectional Long Short-Term Memory (BILSTM); Convolutional Neural Network (CNN); Natural Language Processing (NLP); fake news**

I. Introduction

Nowadays, social media has become an integral part of people's lives [1]. It's a fertile ground for connecting people, creating and sharing information, and staying up to date on trending events [2]. However, these advantages, it's become more difficult for many people to find the difference between confirmed facts and low-quality information with purposefully false data, commonly known as fake news [3]. The increased prevalence of fake news is a worrying phenomenon that has the ability to influence individuals' decisions, opinions and that may lead to severe influence on both people and society [4]. Consequently, an increasing number of researchers are focusing their efforts on identifying dubious information and fake news on online social media platforms and trying to develop effective and automatic systems for detecting online fake news using Artificial Intelligence (AI) techniques.

AI techniques have been successfully used in recent years to solve different prediction and classification problems in large range of research fields [5]–[15].

Specifically, the use of either classic machine learning methods or deep learning techniques in the fake news detection [16] have showed an encouraging results. Traditional machine learning approaches are ineffective for dealing with complicated real-world situations like Natural Language Processing (NLP) tasks and text classification [17], [18]. When compared to Machine Learning (ML) techniques, Deep Learning (DL) has shown significant improvements over time [19]. Furthermore, multiple preprocessing and feature engineering techniques are required for traditional ML algorithms. On the other hand, DL approaches may be able to automatically find useful features in content [20].

In this paper, we present a novel approach to detect fake news based on news content with concatenation of word embedding vectors, and a hybrid approach that combines Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BILSTM). The main contributions of this study are:

- Proposing a novel CNN-BILSTM hybrid approach and extensive experiments are provided to perform the proposed classifier.
- Using concatenation of word embedding technique Word2Vec (CBOW), Word2Vec (Skip-Gram) and Glove.

The rest of this paper is organized as follows: In Section II some important related works are cited. The mathematical problem formulation and the proposed fake news detection system architecture are described in Section III. The used methods of detection and classification are explained in Section IV. Results and discussion are given in Section V. Finally. Section VI concluded this paper.

II. Related Works

In recent years, many researchers have attempted to create a system for identifying the fake news [21]–[28]. In [29], the authors achieved an accuracy of 92%. The authors used a Term Frequency and Inverse Document Frequency (TF-IDF) features-based approach to detect fake news, and only machine learning traditional classification techniques. The study compared six different supervised classifiers, K-Nearest Neighbor (KNN), Linear Support Vector Machine (LSVM), Logistic Regression (LR), Decision Tree (DT), Support

Vector Machine (SVM), and Stochastic Gradient Descent (SGD). The SVM classifier achieved the highest accuracy.

To detect emerging rumors of breaking news, the authors in [30] proposed an approach based on word embedding using Word2Vec and trains a recurrent neural network LSTM. The accuracy achieved was 79.5%. However, the model still requires more improvement. In another study [31], the authors proposed a hybrid method based on an LSTM-CNN model for the classification of tweets into rumors and non-rumors (fake vs genuine). The proposed method achieved an accuracy of 82%.

In [32], the authors used a hybrid method that combines BILSTM with different CNN architectures to detect rumors. They use various pre-trained embedded layers. The best accuracy of the model was 86.12%. The authors of [33] used a hybrid method that combined CNN, LSTM, and BILSTM to develop different models to detect fake news based on the relationship between article headline and article body. The best accuracy of the proposed models was achieved at 71.2%.

The authors in [34] proposed a novel approach based on word embedding over linguistic features by merging linguistic features with word embedding vectors, then using voting classification approach. The best accuracy achieved was 96.73%. In [35], the authors used linguistic features (stylometric, semantic, and syntactic, …). Moreover, for classification they use voting method based on traditional machine learning algorithms. The best accuracy achieved was 96.36%.

III. Proposed Fake News Detection System Architecture

A. Problem Formulation

In this paper, we introduce the problem of fake news detection as shown in Fig. 1.

Given a news content from a corpus of news titles, the objective is to determine if this text content is a fake news. We can formulate this task like a binary classification problem:

Let $T_i = \{w_1 ; w_2 \ldots w_L\}$ be a sequence of words (tokens) of length L of textual content T_i. Given T_i as an input, the first goal is to represent this sequence of words as a sequence of meaningful vectors $Ev_i = \{u_1 ; u_2 \ldots u_L\}$ with $u_i \in \mathbb{R}^D$, the second objective is to train a hybrid CNN-RNN model to classify Ev_i as a fake news or real news by giving a label from the set $Y=\{0$: for fake news ; 1 : for real news$\}$.

B. Proposed Approach

We proposed a four-step framework to detect fake news based on news content. The structure of the proposed system is shown in Fig. 2.

In the first step, we start by applying a series of preprocessing operations on the input data. Among these preprocessing operations, reducing noise, deleting unnecessary repetitions, removing punctuation and alphanumeric characters [36].

Indeed, the preprocessing step is a critical stage in every NLP task. Before we provide the data to the DL model, we cleaned the raw text by removing punctuation. In this step, we used predefined function in python and regular expression to replace all punctuation with empty string. Moreover, we tried to remove all alphanumeric characters such as hashtags and URLs. For every article, we convert all uppercase characters to lowercase. Finally, we split the cleaned article into vectors of words called tokens.

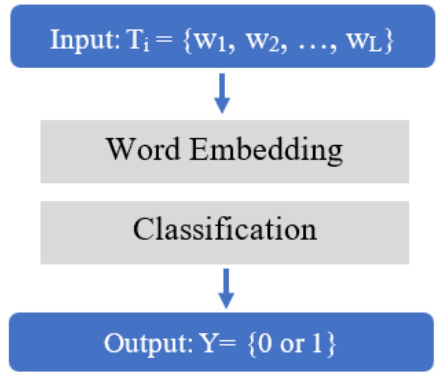


Fig. 1. Formulation of the Proposed Approach of Detection Fake News.

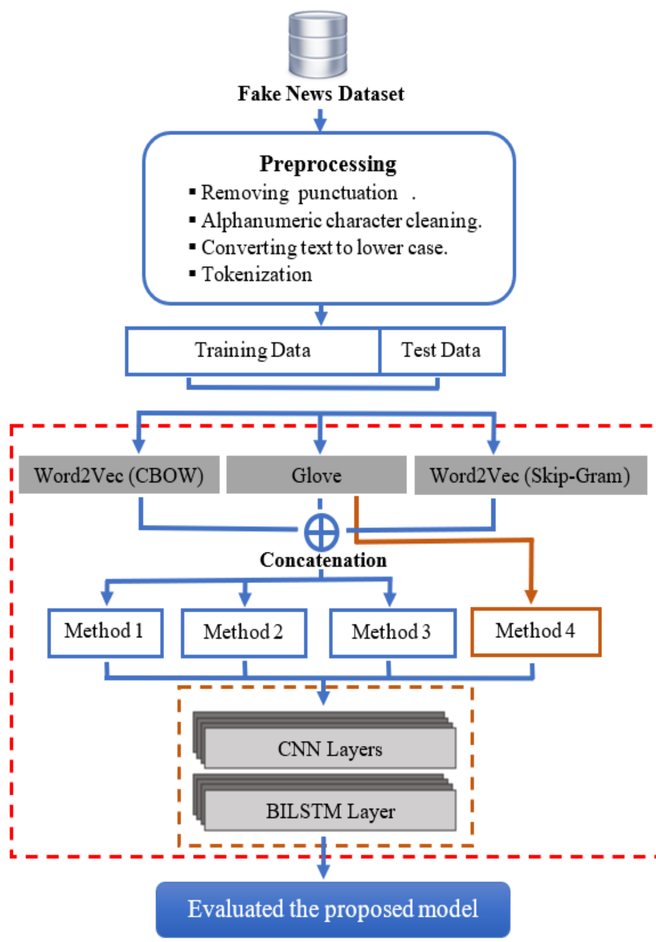


Fig. 2. The Flowchart of the Proposed Approach using CNN-BILSTM Classifier.

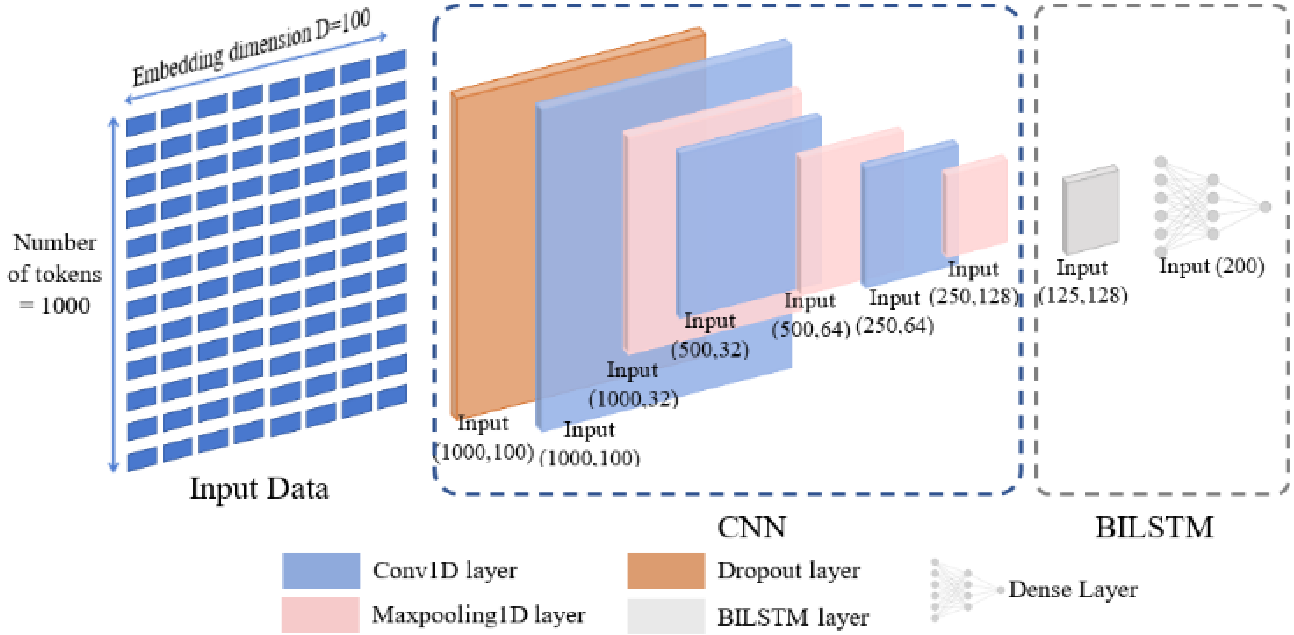


Fig. 3. The Architecture of the Proposed CNN-BILSTM Model

In the next step, the preprocessed input data was converted into word embedding vectors based on the concatenation of two word-embedding techniques. To study the influence of word embedding combination techniques on the proposed model, we proposed four methods to representing words. In method 1, we concatenate the obtained word representation vector using the Word2Vec (CBOW) and the Word2Vec (Skip-Gram) techniques. Method 2 was based on the concatenation of Glove and Word2Vec (Skip-Gram) vectors. In method 3, we concatenate the obtained word representation vector using the Glove and Word2Vec (CBOW) techniques. Method 4 (Baseline) used the Glove word embedding technique. For each method, we provide the obtained vectors to the proposed hybrid CNN-BILSTM network. Then, we conducted four different experiments to study the effect of concatenation. The next step was training the hybrid model. The objective of using CNN layers in this model architecture was to extract fake news features. While, the objective of using BILSTM was to capture the dependencies in the data. The fourth step is to evaluate and compare the application of the four methods on the model.

In Fig. 3, we illustrate CNN-BILSTM architecture used to train the model. From this architecture, we have the convolutional layer that plays a vital role in CNN architecture. It is comprised of a set of convolutional filters [37]. The input data is convolved with these filters to generate the output feature maps [38]. Moreover, the pooling layer is used to sub-sampling and reducing the feature maps dimension with preserving the majority of dominant information. The pooling layer applies statistical methods, such as max pooling, average pooling [39]. The BILSTM Layer. Finally, the Dense Layer consists to connect each unit (neuron) to all neurons of the previous layer and next layer. This layer is activated by an activation function and it is commonly utilized as the CNN classifier [40], [41]. To overcome the overfitting problem, we used regularization technique based on dropout layers by dropping randomly a number of previous layer outputs during training. This technique is efficient for improving the performance of neural networks especially in vision, speech recognition, document classification tasks [42].

IV. Materials and Methods

A. Dataset Description

In this study we use WELFake dataset from [34]. It is one of the largest available datasets that contains 72134 news articles with 37106 fake and 35028 real news. These datasets respect the standards and rules of creating unbiased fake news datasets [43] which in summary are: 1) All articles of dataset should be labeled by experts. 2) Fake news should be collected from different sources. 3) Obtain real news from credible journalism organizations. 4) Collect articles from a variety of news categories in order to create a diverse collection of credible news. The WELFake dataset contains two features (title, content). In addition, all articles are labeled as follow: 0 for fake news and 1 for real news.

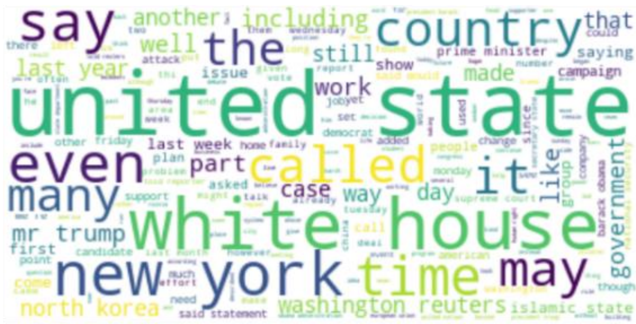
As shown in Table I, the dataset was divided into two parts: The training dataset contains 57229 articles, which represents 80% of the global data. Moreover, the testing and validation dataset represent 20% with 14308 articles.

To have a clear view on the dataset, Fig. 4 illustrates the Word Cloud representation for each class.

TABLE I. Number of Used Fake and Real News Articles

Label	**Train**	**Validation and Test**
Real	29207	7302
Fake	28022	7006
Total	57229	14308

(a)



(b)

Fig. 4. Word cloud Representations of Dataset: (a) for Real News, (b) for Fake News.

B. Word Representation

Word Embedding (WE) is the representation technique of words that converts words to numerical vectors in $\mathbb{R}^D$, whose relative similarities correlate with semantic similarity [44]. Several pretrained word-embedding presentations exist for direct use. The most popular word embedding techniques are:

1) Word2vec: Word2vec is a well-known word embedding algorithm that has two main variants: Skip-Gram and CBOW, both of them are trained by using a neural prediction-based model [45], Skip-Gram predicts the surrounding words (context-words) from the target word [46]. Given a sequence of training words $w_1, w_2, w_3, \ldots, w_L$, Skip-Gram models are trained by maximizing the average log probability:

$$\frac{1}{L}\sum_{t=1}^{L}\sum_{\substack{j=-c,\\ j\neq 0}}^{j=c}\log p\left(w_{t+j}|w_t\right) \tag{1}$$

Where c is the size of the training context, and $\{w\text{-}c, \ldots, w_c\}$ is the word context of the target word wt. In contrast of Skip-Gram, the objective of Continuous Bag-of-Words Model (CBOW) is to predict a word given its context, CBOW models are trained by maximizing the average log probability:

$$\frac{1}{L}\sum_{t=1}^{L}\log\ p\left(w_t|w_{t-c},\ldots,w_{t-1},w_{t+1},\ldots,w_{t+c}\right) \tag{2}$$

2) Glove: Glove is a model that produces a word vector space with meaningful substructure, a weighted least squares model trained on global co-occurrence counts of words and therefore making effective use of statistics [47]. The Glove model is trained by minimizing the loss function

$$\sum_{i=1}^{L}\sum_{j=1}^{L} f\left(X_{ij}\right)\left(v_i^T u_j+b_i+c_j-\log X_{ij}\right)^2 \tag{3}$$

Where $v_i \in \mathbb{R}^D$ is the vector representation of the word w_i, X_{ij} is a weighted count of times word w_j occurs in the context of word w_i, u_j is separate context word vectors, Parameters c_j and b_i represent the bias terms for u_j and v_i, and f is the weighting function defined as follow:

$$f(x)=\min((x/x_{max})^{\alpha};1) \tag{4}$$

There are different embedding vector sizes with D=50, D=100, D= 200, and D=300 dimensions. In this paper we choose the pretrained Glove model that trained on one billion words dataset with a dictionary of 400 thousand tokens and D=50.

C. Deep Learning Methods

1) Convolutional neural network: Convolutional Neural Network (CNN) is a sub-class of deep neural networks. It is particularly used for multimedia content classification tasks such as images and texts, and for time series analysis. One of the main advantages of CNN is recognizing significant features without human intervention [1] that means the ability to solve problems that not contain features which are spatially or temporally dependent. The CNN is inspired from Multi-Layer Perceptron (MLP). Every traditional CNN model includes an input and output layers, and a combination of hidden layers to extract discriminative features [48].

2) LSTM: Long short-term memory (LSTM) is an improved version of recurrent layers applied to learn order dependence in time series and sequence data. LSTMs are similar to RNNs, except that the hidden layer is updated using purpose-built memory cells in order to extract dependencies in the data [49]. Fig. 5 illustrates a LSTM memory cell structure.

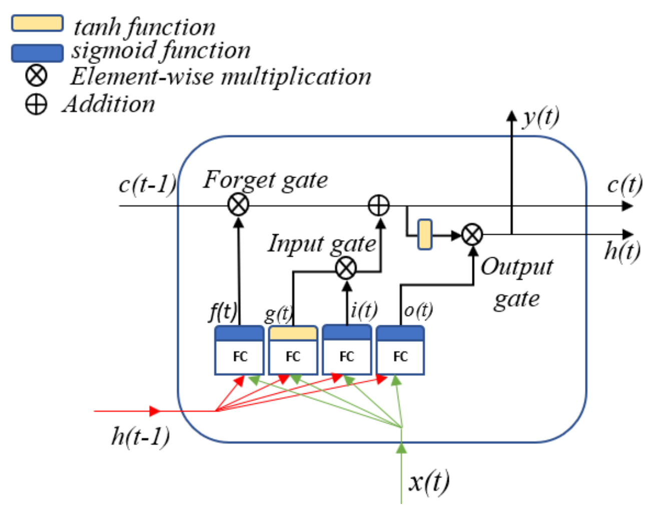


Fig. 5. Single LSTM Memory Cell.

LSTM cell is composed of three gates: input gate, forget gate and output gate. The functions of gates are defined as follow:

$$f_{(t)}=\sigma\left(W_{xf}^{T}x_{(t)}+W_{hf}^{T}h_{(t-1)}+b_f\right) \quad (5)$$

$$i_{(t)}=\sigma\left(W_{xi}^{T}x_{(t)}+W_{hi}^{T}h_{(t-1)}+b_i\right) \quad (6)$$

$$g_{(t)}=\tanh\left(W_{xg}^{T}x_{(t)}+W_{hg}^{T}h_{(t-1)}+b_g\right) \quad (7)$$

$$c_{(t)}=f_{(t)}\otimes c_{(t-1)}+i_{(t)}\otimes g_{(t)} \quad (8)$$

$$y_{(t)}=h_t=o_{(t)}\otimes\tanh\left(c_{(t)}\right) \quad (9)$$

$$o_{(t)}=\sigma\left(W_{xo}^{T}x_{(t)}+W_{ho}^{T}h_{(t-1)}+b_o\right) \quad (10)$$

Where, and W_x and W_h are the weight matrices and b is the bias.

3) BILSTM: The BILSTM neural network is an advanced version of LSTM. The BILSTM architecture contains LSTM units that function in both directions to integrate left and right context information. Bi-directional network consists of two parallel layers: The forward layer and backward layer extract representation of the left and right context. The outputs of these two layers are then concatenated to produce a complete representation of an element of the input sequence [50]. Fig. 6 illustrates the BILSTM network architecture.

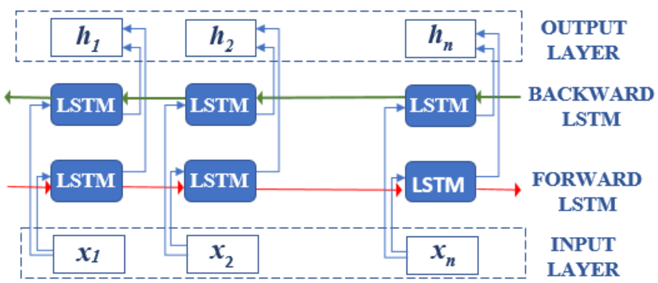


Fig. 6. Simple BILSTM Network Architecture.

V. Results and Discussion

A. Experimental Setup and Optimal Algorithms Configuration

In this section, we present the experiment and evaluation results on fake news detection. The experimental models are implemented inside an open cloud environment Google Colab that provides a Tesla K80 GPU with 12GB of GDDR5 VRAM. Python 3.7 and different libraries like TensorFlow and Scikit-learn are used to build the models. In order to find the optimum hyper-parameters of the proposed model, the experiments were conducted with grid search. Table II shows the optimal parameters setting of the model architecture.

B. Evaluation Metrics

We use four metrics to evaluate the results based on the number of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) in the binary classification:

$$\text{Accuracy}=\frac{TP+TN}{TP+FP+TN+FN} \quad (11)$$

$$\text{Recall}=\frac{TP}{TP+FN} \quad (12)$$

$$F_1\text{ score}=\frac{2*(\text{precision}\times\text{recall})}{\text{precision}+\text{recall}} \quad (13)$$

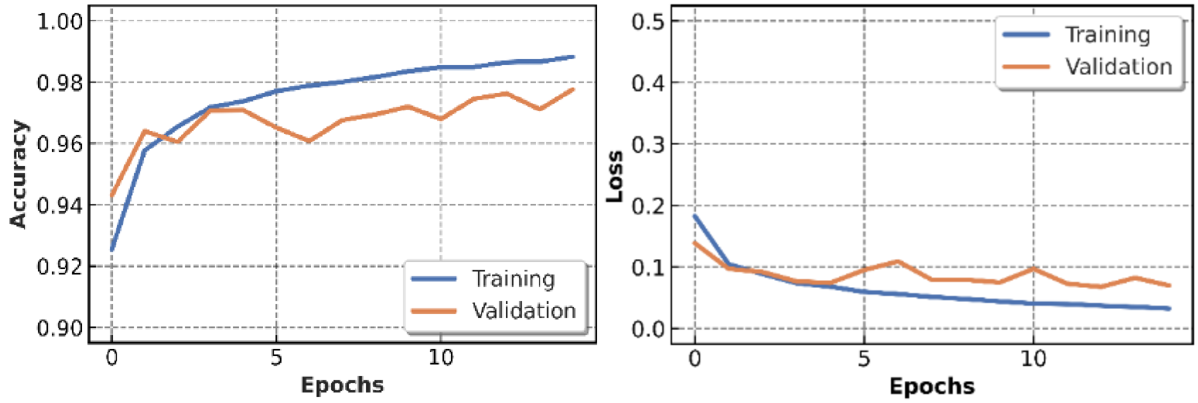
$$\text{Precision}=\frac{TP}{TP+FP} \quad (14)$$

C. Training Results

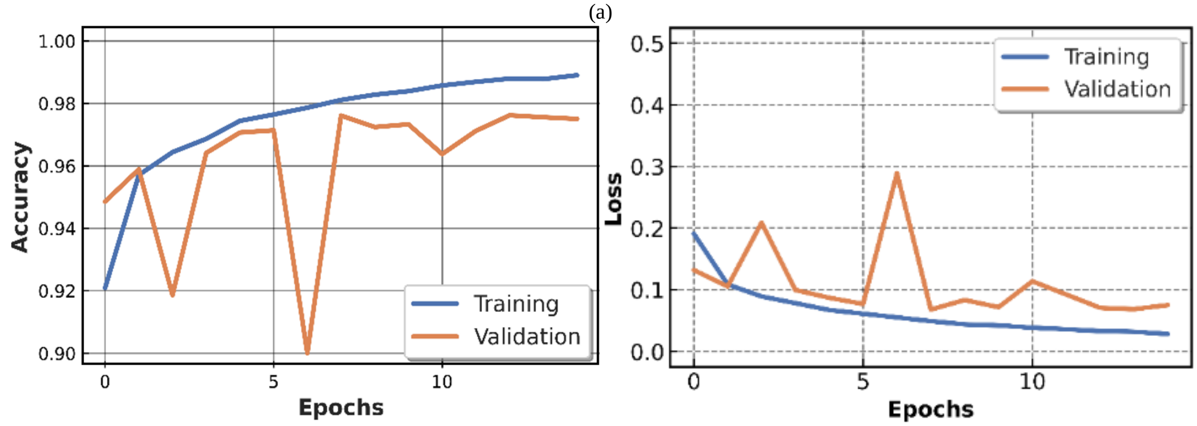
Fig. 7 shows the variation of accuracy and loss metrics across epochs. On training, the improvement in term of accuracy and loss of the three models is linear until the 10th epoch. After that, the training performance begins to stabilize and the improvement becomes slow. On the opposite, the validation performance increase with unstable manner.

TABLE II. Parameters Settings of the BILSTM-CNN Classifier

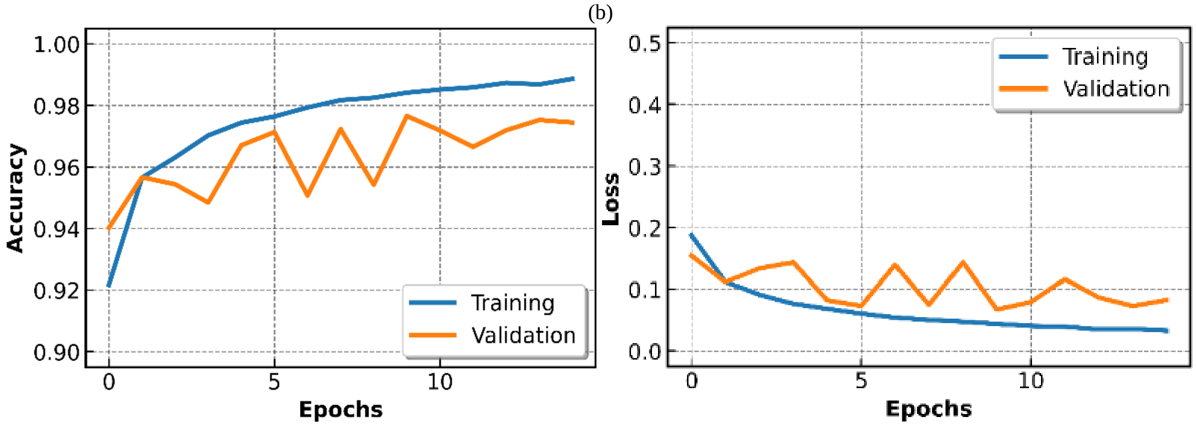
Parameter	**Value**
Vocabulary size	20 000
Input vector	1000
Word Embedded size	100
Number of convolutional layers	3
Number of filters	128
Filter size	3,5,7
Dropout rate	0,2
Activation function	Sigmoid
Number of epochs	15
Optimization	Adam



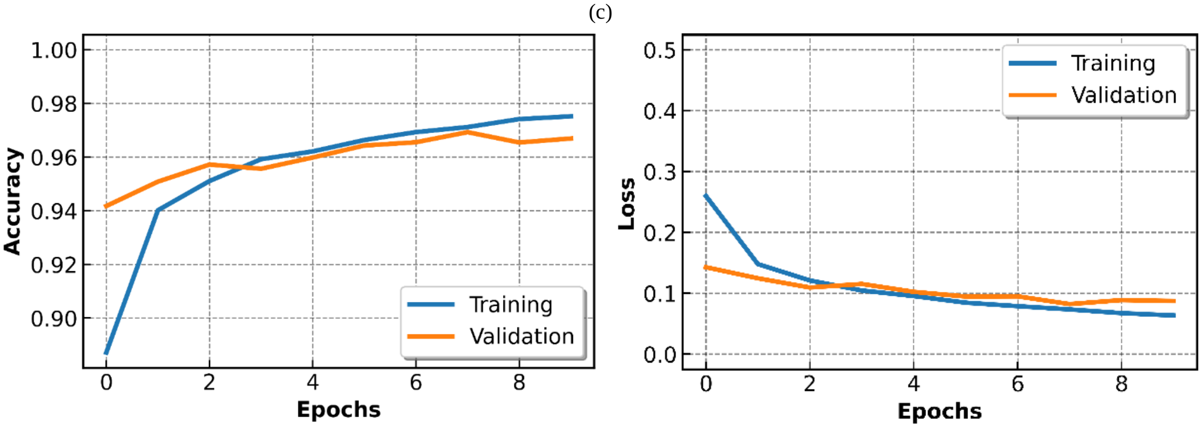
(a)



(b)



(c)



(d)

Fig. 7. Accuracy and Loss Curves Obtained by Training and Validating of Proposed Models. (a) Method 1, (b) Method 2, (c) Method 3 and (d) Method 4 (Baseline).

D. Testing Results

Training a CNN models provides a prediction-ready model with the appropriate weights that corresponds to the task being performed. As indicated in Table III, a comparative result in terms of accuracy, recall, F1-score and precision of the model variants. We can clearly observe that the combined word embedded techniques reached an improvement when compared to baseline method.

Among all methods, the method 1 provided the best result in terms of accuracy (97.74%) and precision (97.74%). While, the method 4 based on one-word embedding technique is the lowest in term of accuracy, F1-score and Precision. The method 1 shows better performance in the detection of fake news. The confusion matrix of the proposed model is shown in Fig. 8. We found 3472 news was predicted as fake and was actually fake news. In addition, 66 news was predicted as fake was actually real. On the other hand, 3520 real news were predicted as real and was actually real. Moreover, 96 was predicted as real and was actually fake.

To investigate these findings, a ROC curve illustrating the sensitivity vs specificity (or False Positive Rate versus True Positive Rate) of a diagnostic test might be employed. This shape of the curve enables us to compare several models depending on the value of the AUC variable. This number indicates the total area under the ROC curve between two dimensions. This article depicts the ROC curve for each model considered in this investigation. Fig. 9 depicts a scatterplot of the False Positive Rate (TPR) vs the True Positive Rate (TNR) for the proposed CNN-BILSTM model using word embedding Method 1.

TABLE III. MODELS PERFORMANCE WITH DIFFERENT COMBINATION OF WORD EMBEDDING TECHNIQUES

	Accuracy	Recall	F1-score	Precision
Method 1	**97.74%**	97,35%	97,75%	**98,16%**
Method 2	97,71%	98,59%	97,75%	96,93%
Method 3	97,46%	96,99%	97,47%	97,96%
Method 4 (Baseline)	96,56%	97,54%	96,63%	95,74%

Fig. 8. Confusion Matrix for CNN (Method 1).

E. Discussion

In this study, we proposed a hybrid deep learning model to detect fake news based on news content. we used WELFake dataset which is one of the biggest databases accessible and comprises 72134 news, of which 37106 are fraudulent and 35028 are authentic. These datasets adhere to the principles and guidelines of developing impartial fake news. The WELFake dataset includes two characteristics (title, content). Additionally, each article is categorized as follows: zero for false news and one for genuine news. In the first step, we started by applying a series of preprocessing operations on the input data. These operations include reducing noise, deleting unnecessary repetitions, removing punctuation and alphanumeric characters. The preprocessed input data was converted into word embedding vectors based on the concatenation of two word-embedding techniques.

In proposed system architecture, we presented four approaches to represent words in order to examine the effect of word embedding combination strategies on proposed model. In approach 1, the acquired word representation vector is concatenated using Word2Vec (CBOW) and Word2Vec (Skip-Gram). The second method was based on the union of the Glove and Word2Vec (Skip-Gram) vectors. The acquired word representation vector is concatenated using the Glove and Word2Vec (CBOW) approaches in method 3. Method 4 (Baseline) used the Glove approach for word embedding. We submit the acquired vectors for each approach to the suggested hybrid CNN-BILSTM network. For each method, we provided the obtained vectors to the proposed CNN-BILSTM network. The objective of using CNN layers in model architecture was to extract fake news features. Then, we provided the acquired vectors for each approach to the suggested CNN-BILSTM network. The goal of using CNN layers into the model architecture was to extract elements of fake news. The experiments concluded that the combination of two-word embedding vectors offers a benefit and it increases the accuracy of the proposed model when compared to using one-word embedding technique. Table III and Fig. 11 show that the combination of Word2Vec (CBOW) and the Word2Vec (Skip-Gram) gives the best results in term of accuracy and precision when compared to the other possible combinations.

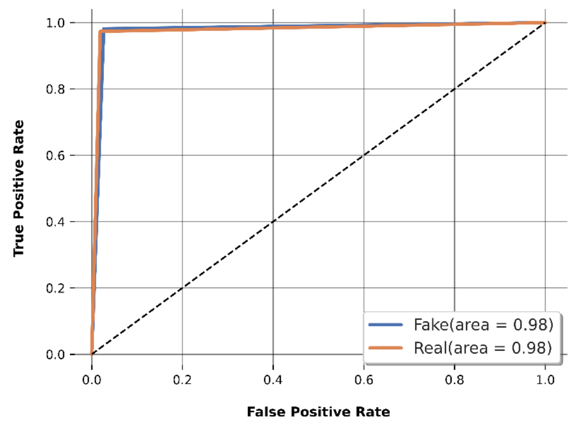


Fig. 9. The ROC Curves of Method 1.

Overall, the results of the experiment indicate that all models attained exceptional precision with a low loss rate. In addition, these findings demonstrate that the incorporation of a hybrid model based on the feature descriptor approach improves prediction performance relative to other models in the literature.

In Table IV, we compare the performance of the proposed model to that of some similar research examined in the present paper.

The Fig. 10 shows the improvement of the proposed method 1 compared to the traditional machine learning methods. The accuracy of the proposed model achieved 97.74%, while DT, RF, LR and SVM achieved a maximum accuracy only up to 93.51%, 94.37%, 95.42% and 97.09%, respectively.

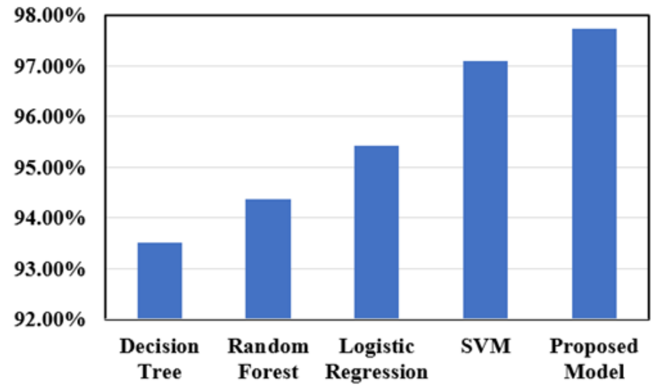


Fig. 10. Accuracy Comparison of Proposed Model with Traditional Machine Learning Algorithms.

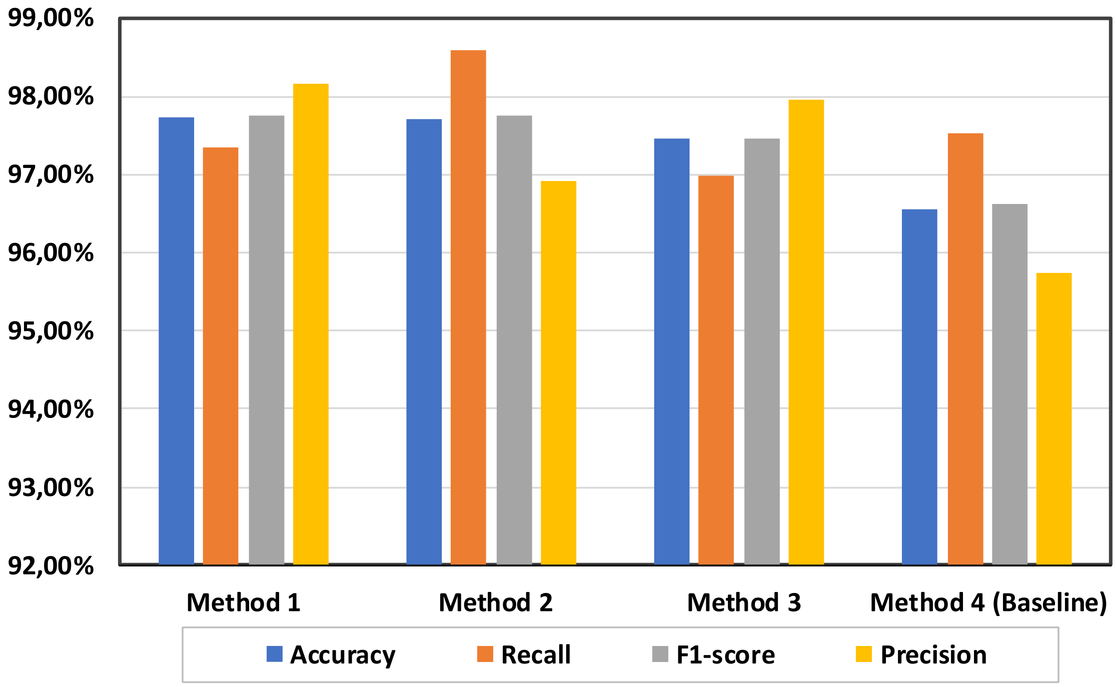


Fig. 11. Summarize of All the Experimental Findings Provided in this Research.

TABLE IV. ACCURACY COMPARISON OF PROPOSED MODEL WITH RELATED WORKS

Authors	**Dataset**	**Feature Representation**	**Classification Algorithm**	**Accuracy**
[29]	Their own dataset of 25200 articles from Reuters and Kaggle	TF-IDF	KNN, SVM, LR, LSVM, DT	92%
[30]	PHEME	WORD2VEC	LSTM	79.5%
[31]	PHEME	WORD2VEC	Hybrid LSTM-CNN	82%
[32]	PHEME	WORD2VEC-GLOVE-FASTTEXT	Hybrid BILSTM-CNN	86.12%
[33]	FNC-1	GLOVE	Hybrid LSTM/ BILSTM -CNN	71.2%
[35]	ISOT	Linguistic features (stylometric, semantic, and syntactic…)	Voting classifier based on traditional machine learning algorithms	96.36%
[34]	WELFake dataset	Combine linguistic feature with Word embedding	Voting classifier based on SVM , DT , NB, Bagging, AdaBoost, KNN	96.73%
Proposed method	**WELFake dataset**	**Combine linguistic feature with Word embedding**	**CNN-BILSTM**	**97.74%**

VI. CONCLUSION AND PERSPECTIVES

In this paper, we presented a new approach to detect fake news using the unbiased dataset WELFake. The main objective of this research was obtaining a good result and improving the performance of the proposed fake news detection system. In this study. The first task of this approach was representing words to meaningful numerical vectors using combination of different word embedding techniques. The second task was training the proposed hybrid model based on CNN and BILSTM architectures. The obtained results show an improvement in terms of accuracy and precision when compared to traditional machine learning algorithms and related work results. The simple concatenation of the different pre-trained embedding models increases the dimension of embedded vectors. Moreover, that leads to a high computational complexity. To solve this problem in future works, we propose to reduce the dimension of concatenated vectors with preserving characteristics of the original data using features engineering techniques such as features selection and dimensions reduction.

REFERENCES

[1] J. Gu et al., 'Recent Advances in Convolutional Neural Networks', Pattern Recognition, vol. 77, Dec. 2015, doi: 10.1016/j.patcog.2017.10.013.

[2] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, 'Fake News Detection on Social Media: A Data Mining Perspective', arXiv:1708.01967 [cs], Sep. 2017, Accessed: Apr. 07, 2022. [Online]. Available: http://arxiv.org/abs/1708.01967.

[3] F. B. Gereme and W. Zhu, 'Early Detection of Fake News "Before It Flies High"', in Proceedings of the 2nd International Conference on Big Data Technologies - ICBDT2019, Jinan, China, 2019, pp. 142–148. doi: 10.1145/3358528.3358567.

[4] D. M. J. Lazer et al., 'The science of fake news', Science, vol. 359, no. 6380, pp. 1094–1096, Mar. 2018, doi: 10.1126/science.aao2998.

[5] S. Hamida, B. Cherradi, and H. Ouajji, 'Handwritten Arabic Words Recognition System Based on HOG and Gabor Filter Descriptors', in 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, Apr. 2020, pp. 1–4. doi: 10.1109/IRASET48871.2020.9092067.

[6] O. Terrada, S. Hamida, B. Cherradi, A. Raihani, and O. Bouattane, 'Supervised Machine Learning Based Medical Diagnosis Support System for Prediction of Patients with Heart Disease', Adv. sci. technol. eng. syst. j., vol. 5, no. 5, pp. 269–277, 2020, doi: 10.25046/aj050533.

[7] H. Moujahid, B. Cherradi, O. E. Gannour, L. Bahatti, O. Terrada, and S. Hamida, 'Convolutional Neural Network Based Classification of Patients with Pneumonia using X-ray Lung Images', Adv. sci. technol. eng. syst. j., vol. 5, no. 5, pp. 167–175, 2020, doi: 10.25046/aj050522.

[8] O. El Gannour, S. Hamida, B. Cherradi, A. Raihani, and H. Moujahid, 'Performance Evaluation of Transfer Learning Technique for Automatic Detection of Patients with COVID-19 on X-Ray Images', in 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra, Morocco, Dec. 2020, pp. 1–6. doi: 10.1109/ICECOCS50124.2020.9314458.

[9] O. Terrada, A. Raihani, O. Bouattane, and B. Cherradi, 'Fuzzy cardiovascular diagnosis system using clinical data', in 2018 4th International Conference on Optimization and Applications (ICOA), Mohammedia, Apr. 2018, pp. 1–4. doi: 10.1109/ICOA.2018.8370549.

[10] B. Cherradi, O. Terrada, A. Ouhmida, S. Hamida, A. Raihani, and O. Bouattane, 'Computer-Aided Diagnosis System for Early Prediction of Atherosclerosis using Machine Learning and K-fold cross-validation', in 2021 International Congress of Advanced Technology and Engineering (ICOTEN), Taiz, Yemen, Jul. 2021, pp. 1–9. doi: 10.1109/ICOTEN52080.2021.9493524.

[11] S. Laghmati, B. Cherradi, A. Tmiri, O. Daanouni, and S. Hamida, 'Classification of Patients with Breast Cancer using Neighbourhood Component Analysis and Supervised Machine Learning Techniques', in 2020 3rd International Conference on Advanced Communication Technologies and Networking (CommNet), Marrakech, Morocco, Sep. 2020, pp. 1–6. doi: 10.1109/CommNet49926.2020.9199633.

[12] O. Terrada, B. Cherradi, A. Raihani, and O. Bouattane, 'Atherosclerosis disease prediction using Supervised Machine Learning Techniques', in 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, Apr. 2020, pp. 1–5. doi: 10.1109/IRASET48871.2020.9092082.

[13] O. Daanouni, B. Cherradi, and A. Tmiri, 'Predicting diabetes diseases using mixed data and supervised machine learning algorithms', in Proceedings of the 4th International Conference on Smart City Applications - SCA '19, Casablanca, Morocco, 2019, pp. 1–6. doi: 10.1145/3368756.3369072.

[14] O. Terrada, B. Cherradi, S. Hamida, A. Raihani, H. Moujahid, and O. Bouattane, 'Prediction of Patients with Heart Disease using Artificial Neural Network and Adaptive Boosting techniques', in 2020 3rd International Conference on Advanced Communication Technologies and Networking (CommNet), Marrakech, Morocco, Sep. 2020, pp. 1–6. doi: 10.1109/CommNet49926.2020.9199620.

[15] S. Hamida, B. Cherradi, O. Terrada, A. Raihani, H. Ouajji, and S. Laghmati, 'A Novel Feature Extraction System for Cursive Word Vocabulary Recognition using Local Features Descriptors and Gabor Filter', in 2020 3rd International Conference on Advanced Communication Technologies and Networking (CommNet), Marrakech, Morocco, Sep. 2020, pp. 1–7. doi: 10.1109/CommNet49926.2020.9199642.

[16] X. Zhang and A. A. Ghorbani, 'An overview of online fake news: Characterization, detection, and discussion', Information Processing & Management, vol. 57, no. 2, p. 102025, Mar. 2020, doi: 10.1016/j.ipm.2019.03.004.

[17] S. Hamida, B. Cherradi, and H. Ouajji, 'Handwritten Arabic Words Recognition System Based on HOG and Gabor Filter Descriptors', in 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, Apr. 2020, pp. 1–4. doi: 10.1109/IRASET48871.2020.9092067.

[18] S. Hamida, B. Cherradi, O. Terrada, A. Raihani, H. Ouajji, and S. Laghmati, 'A Novel Feature Extraction System for Cursive Word Vocabulary Recognition using Local Features Descriptors and Gabor Filter', in 2020 3rd International Conference on Advanced Communication Technologies and Networking (CommNet), Marrakech, Morocco, Sep. 2020, pp. 1–7. doi: 10.1109/CommNet49926.2020.9199642.

[19] Y. LeCun, Y. Bengio, and G. Hinton, 'Deep learning', Nature, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[20] M. Al-Sarem, A. Alsaeedi, F. Saeed, W. Boulila, and O. AmeerBakhsh, 'A Novel Hybrid Deep Learning Model for Detecting COVID-19-Related Rumors on Social Media Based on LSTM and Concatenated Parallel CNNs', Applied Sciences, vol. 11, no. 17, p. 7940, Aug. 2021, doi: 10.3390/app11177940.

[21] F. L. Alotaibi and M. M. Alhammad, 'Using a Rule-based Model to Detect Arabic Fake News Propagation during Covid-19', IJACSA, vol. 13, no. 1, 2022, doi: 10.14569/IJACSA.2022.0130114.

[22] Y. Alsaawy, A. Alkhodre, N. M. Bahbouh, A. A. Sen, and A. Nadeem, 'Lightweight Chain for Detection of Rumors and Fake News in Social Media', IJACSA, vol. 12, no. 8, 2021, doi: 10.14569/IJACSA.2021.0120860.

[23] K.-H. Kim and C.-S. Jeong, 'Fake News Detection System using Article Abstraction', in 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE), Chonburi, Thailand, Jul. 2019, pp. 209–212. doi: 10.1109/JCSSE.2019.8864154.

[24] X. Cui and Y. Li, 'Fake News Detection in Social Media based on Multi-Modal Multi-Task Learning', IJACSA, vol. 13, no. 7, 2022, doi: 10.14569/IJACSA.2022.01307106.

[25] A. R. Mahlous and A. Al-Laith, 'Fake News Detection in Arabic Tweets during the COVID-19 Pandemic', IJACSA, vol. 12, no. 6, 2021, doi: 10.14569/IJACSA.2021.0120691.

[26] S. Raza and C. Ding, 'Fake news detection based on news content and social contexts: a transformer-based approach', Int J Data Sci Anal, vol. 13, no. 4, pp. 335–362, May 2022, doi: 10.1007/s41060-021-00302-z.

[27] M. Aldwairi and A. Alwahedi, 'Detecting Fake News in Social Media Networks', Procedia Computer Science, vol. 141, pp. 215–222, 2018, doi: 10.1016/j.procs.2018.10.171.

[28] A. Albahr and M. Albahar, 'An Empirical Comparison of Fake News Detection using different Machine Learning Algorithms', IJACSA, vol. 11, no. 9, 2020, doi: 10.14569/IJACSA.2020.0110917.

[29] H. Ahmed, I. Traore, and S. Saad, 'Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques', in Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, vol. 10618, I. Traore, I. Woungang, and A. Awad, Eds. Cham: Springer International Publishing, 2017, pp. 127–138. doi: 10.1007/978-3-319-69155-8_9.

[30] S. A. Alkhodair, S. H. H. Ding, B. C. M. Fung, and J. Liu, 'Detecting breaking news rumors of emerging topics in social media', Information Processing & Management, vol. 57, no. 2, p. 102018, Mar. 2020, doi: 10.1016/j.ipm.2019.02.016.

[31] O. Ajao, D. Bhowmik, and S. Zargari, 'Fake News Identification on Twitter with Hybrid CNN and RNN Models', 2018, doi: 10.48550/ARXIV.1806.11316.

[32] M. Z. Asghar, A. Habib, A. Habib, A. Khan, R. Ali, and A. Khattak, 'Exploring deep neural networks for rumor detection', J Ambient Intell Human Comput, vol. 12, no. 4, pp. 4315–4333, Apr. 2021, doi: 10.1007/s12652-019-01527-4.

[33] A. Abedalla, A. Al-Sadi, and M. Abdullah, 'A Closer Look at Fake News Detection: A Deep Learning Perspective', in Proceedings of the 2019 3rd International Conference on Advances in Artificial Intelligence, Istanbul Turkey, Oct. 2019, pp. 24–28. doi: 10.1145/3369114.3369149.

[34] P. K. Verma, P. Agrawal, I. Amorim, and R. Prodan, 'WELFake: Word Embedding Over Linguistic Features for Fake News Detection', IEEE Trans. Comput. Soc. Syst., vol. 8, no. 4, pp. 881–893, Aug. 2021, doi: 10.1109/TCSS.2021.3068519.

[35] J. F. Ribeiro Bezerra, 'Content-based fake news classification through modified voting ensemble', Journal of Information and Telecommunication, vol. 5, no. 4, pp. 499–513, Oct. 2021, doi: 10.1080/24751839.2021.1963912.

[36] S. Hamida, B. Cherradi, O. El Gannour, O. Terrada, A. Raihani, and H. Ouajji, 'New Database of French Computer Science Words Handwritten Vocabulary', in 2021 International Congress of Advanced Technology and Engineering (ICOTEN), Taiz, Yemen, Jul. 2021, pp. 1–5. doi: 10.1109/ICOTEN52080.2021.9493438.

[37] A. Severyn and A. Moschitti, 'Learning to Rank Short Text Pairs with Convolutional Deep Neural Networks', in Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago Chile, Aug. 2015, pp. 373–382. doi: 10.1145/2766462.2767738.

[38] A. Fesseha, S. Xiong, E. D. Emiru, M. Diallo, and A. Dahou, 'Text Classification Based on Convolutional Neural Networks and Word Embedding for Low-Resource Languages: Tigrinya', Information, vol. 12, no. 2, p. 52, Jan. 2021, doi: 10.3390/info12020052.

[39] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, 'l', 2014, doi: 10.48550/ARXIV.1404.2188.

[40] S. Hamida, B. Cherradi, H. Ouajji, and A. Raihani, 'Convolutional Neural Network Architecture for Offline Handwritten Characters Recognition', in Innovation in Information Systems and Technologies to Support Learning Research, vol. 7, M. Serrhini, C. Silva, and S. Aljahdali, Eds. Cham: Springer International Publishing, 2020, pp. 368–377. doi: 10.1007/978-3-030-36778-7_41.

[41] S. Hamida, B. Cherradi, A. Raihani, and H. Ouajji, 'Performance Evaluation of Machine Learning Algorithms in Handwritten Digits Recognition', in 2019 1st International Conference on Smart Systems and Data Science (ICSSD), Rabat, Morocco, Oct. 2019, pp. 1–6. doi: 10.1109/ICSSD47982.2019.9003052.

[42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, 'Dropout: A Simple Way to Prevent Neural Networks from Overfitting', J. Mach. Learn. Res., vol. 15, no. 1, pp. 1929–1958, Jan. 2014.

[43] G. Gravanis, A. Vakali, K. Diamantaras, and P. Karadais, 'Behind the cues: A benchmarking study for fake news detection', Expert Systems with Applications, vol. 128, pp. 201–213, Aug. 2019, doi: 10.1016/j.eswa.2019.03.036.

[44] A. Mandelbaum and A. Shalev, 'Word Embeddings and Their Use In Sentence Classification Tasks', arXiv:1610.08229 [cs], Oct. 2016, Accessed: Apr. 07, 2022. [Online]. Available: http://arxiv.org/abs/1610.08229.

[45] T. Mikolov, K. Chen, G. Corrado, and J. Dean, 'Efficient Estimation of Word Representations in Vector Space', 2013, doi: 10.48550/ARXIV.1301.3781.

[46] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, 'Distributed Representations of Words and Phrases and their Compositionality', 2013, doi: 10.48550/ARXIV.1310.4546.

[47] J. Pennington, R. Socher, and C. Manning, 'Glove: Global Vectors for Word Representation', in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 2014, pp. 1532–1543. doi: 10.3115/v1/D14-1162.

[48] J. Gu et al., 'Recent advances in convolutional neural networks', Pattern Recognition, vol. 77, pp. 354–377, May 2018, doi: 10.1016/j.patcog.2017.10.013.

[49] Z. Huang, W. Xu, and K. Yu, 'Bidirectional LSTM-CRF Models for Sequence Tagging', 2015, doi: 10.48550/ARXIV.1508.01991.

[50] T. A. Le, M. Y. Arkhipov, and M. S. Burtsev, 'Application of a Hybrid Bi-LSTM-CRF Model to the Task of Russian Named Entity Recognition', in Artificial Intelligence and Natural Language, vol. 789, A. Filchenkov, L. Pivovarova, and J. Žižka, Eds. Cham: Springer International Publishing, 2018, pp. 91–103. doi: 10.1007/978-3-319-71746-3_8.