Systems Assessment and Engineering Management

Journal Homepage: sciencesforce.com/saem

SCIENCES FORCE
International Academic Publisher

Syst. Assess. Eng. Manage. Vol. 1 (2024) 45–53

Paper Type: Original Article

Differentiating Authentic News from Fabrications using Deep Learning: A New Approach

Ahmed Elmasry [1,*] **and Nourhan Talal** [1]

[1] Department of Computer Science, Faculty of Computers and Informatics, Zagazig University, Zagazig 44519, Egypt. Emails: a.elmasry24@fci.zu.edu.eg; N.Talal22@fci.zu.edu.eg.

Received: 17 Dec 2023 **Revised:** 13 Mar 2024 **Accepted:** 12 Apr 2024 **Published:** 15 Apr 2024

Abstract

Fake news on social media disrupts perceptions and influences decision-making. Despite the importance of detection in politics, there's a lack of research on AI and ML-focused detection models. In this paper, we suggest a novel hybrid deep learning model (CNN-DNN) that combines the convolution neural network (CNN) and deep neural network (DNN). The proposed model was evaluated on the WELFake dataset, which consists of a total of 72,134 news articles. Out of the total number of articles, 35,028 are categorized as genuine news, while 37,106 are categorized as fabricated news. The proposed model achieved the highest accuracy at 0.9732, however, the LSTM model attained the lowest accuracy at 0.960. Our goal was to develop a detection model that would efficiently curb the dissemination of false information. The efficacy of this model, which depends on multiple data sources, has been demonstrated in the context of managerial decision-making. The study offers practical insights and indicates potential areas for future research.

Keywords: Fake News Classification, Convolution Neural Network, Deep Neural Network.

1 | Introduction

Currently, digital platforms such as social media, online forums, and websites have become more important than traditional media in terms of being the main sources of information [1]. This shift highlights the transformation in our methods of obtaining and interacting with information. The spread of social network platforms exacerbates the problems of fake news (FKs) and disinformation. Hence, it is imperative to employ Artificial Intelligence (AI) as a means to counteract the spread of false information and minimize its detrimental effects.

As reported by CNN In 2021, TikTok became flooded with videos featuring Tom Cruise engaging in activities that were unexpectedly different from his usual demeanor. These activities included playfully behaving in an elegant men's clothing store and demonstrating a coin trick. Considerable focus has been directed towards the possibility of utilizing deep fakes for malicious intentions and with valid justification.

The utilization of AI in the classification of FKs is of utmost importance due to its ability to systematically analyze the intricate nuances of language and contextual information that could be overlooked by humans [2]. The advancements in AI and Natural Language Processing (NLP) have sparked a greater interest in the

Corresponding Author: a.elmasry24@fci.zu.edu.eg

https://doi.org/10.61356/j.saem.2024.1272

classification of FKs. As a result, numerous innovative approaches have been developed for research in this field.

With the ongoing prevalence of misinformation on social media, the credibility of online information is being increasingly questioned. to devise efficient strategies to tackle this issue. As the amount of data continues to grow, it becomes more and more crucial to collect relevant information quickly and effectively. This highlights the significance of employing computational linguistic techniques. Within this context, the utilization of AI methodologies becomes essential, offering sophisticated instruments to accurately identify and tackle misinformation.

In this paper, a novel hybrid DL model is introduced to effectively classify FKs. The model (CNN-DNN) was evaluated on the WELFake dataset, which contains 72,134 news articles, containing 35,028 real news, and 37,106 FKs. Model results according to accuracy, precision, recall, and F1-score have been compared with other DL models. The proposed model achieved the highest accuracy at 0.9732, while the LSTM model attained the lowest accuracy at 0.960.

Therefore, the main contributions of this work are as follows:

- We introduce a novel hybridization between CNN and DNN (CNN-DNN) for FK detection.
- A large scale of FK data has been used for training and testing.
- The proposed model achieves higher performance with 0.973, 0.973, 0.973, and 0.973 in terms of accuracy, precision, recall, and F1-score.
- Utilize NLP methods for text analysis and train deep learning (DL) models to identify false information by analyzing the news title or content.

The rest of this paper is organized as follows: Section 2 reviews the related work. Section 3 presents Materials and methodology including the details of the proposed work. Section 4 is dedicated to presenting the results and discussions. Section 5 presents the Managerial implications. Section 6 concludes the paper and outlines future work.

2 | Related Work

FKs is a burgeoning subject that has garnered the interest of researchers worldwide in the realm of AI. Despite garnering considerable attention in the research community, the lack of context-specific news data has hindered significant improvements in the accuracy of FKs classification. Unlike the traditional feature-based model, DL offers an advantage by eliminating the need for manually designed features. Instead, it autonomously identifies the most suitable feature set for a given classification task or problem.

Palani et. al [3] introduced a novel automated technique for detecting counterfeit news that generates a multimodal feature vector by combining textual and visual data. The CapsNet model beats CNN in extracting visual cues, making it easier to distinguish between real and counterfeit news. The system's performance was evaluated using publicly available datasets, and it achieved 93% and 92% accuracy rates, respectively, compared to 84.6% and 85.6% for the SpotFake+ model.

Sastrawan et. al [4] proposed a DL methodology that incorporates multiple architectures such as Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (LSTM), and Residual Network (ResNet). These architectures are augmented with pre-trained word embedding and trained on four separate datasets. To address imbalances in data distribution across different classes, the authors employ the back-translation method to augment every piece of data.

Zhang et. al [5] proposed advanced DL techniques that are being developed for identifying false information in cyber-physical social services. The model uses Chinese text as the primary unit for processing, extracting feature representation using a convolution-based neural computing framework. Experiments show reduced training time and superior classification accuracy compared to baseline methods.

Mohawesh et. al [6] introduced a semantic methodology for classifying FKs by analyzing relational variables such as sentiment, entities, and facts that can be directly extracted from the text. This model achieved superior performance compared to the most advanced techniques by approximately 3.97% for English to English, 1.41% for English to Hindi, 5.47% for English to Indonesian, 2.18% for English to Swahili, and 2.88% for English to Vietnamese language reviews on the TALLIP FKs dataset.

Soga et. al [7] introduced a detection method that considers the similarity of opinions between users by analyzing their attitudes toward news articles and their interactions with other users' posts. This method employs a graph transformer network to extract both global structural information and interactions among similar stances simultaneously. Moreover, it tackles the difficulties of analyzing viewpoints in microblogs while reducing the impact of inadequately represented stance characteristics. Assessed methodology utilizing specifically gathered Twitter data and the standardized FibVID dataset. The results showed a significant enhancement in the ability to detect compared to traditional approaches.

Abualigah et. al [8] suggested a technique for enhancing the false news detection system. The preprocessing stage, which utilizes Glove, an unsupervised learning algorithm developed by researchers at Stanford University, is where the most notable improvement occurs. Glove generates word embeddings by combining global word co-occurrence matrices from a given corpus. The fundamental concept underlying the GloVe word embedding is to deduce the correlation between words based on statistical analysis. This method incorporates advanced DL algorithms, including CNN, DNN, and long short-term memory (LSTM). The Recurrent Neural Network (RNN) utilizes the GloVe embedding technique during the preprocessing stage. It leverages the Curpos FKs dataset to improve the system's performance. The RNN's sequential processes and classification contribute to achieving an impressive accuracy rate of 98.974%.

3 | Materials and Methodology

3.1. | Dataset

This study utilized the WELFake dataset [9], which comprises 72,134 news articles. Among these articles, 35,028 are classified as real news and 37,106 are classified as FKs. The dataset was compiled from diverse sources to mitigate the issues of overfitting. The dataset comprises two distinct attributes: the title and the content, along with the label, which classifies the dataset as either 0 for FKs or 1 for real news.

3.2. | Dataset Preprocessing

The dataset comprises two features, namely title, and text. To begin processing data, the initial step is to eliminate the rows that lack data. Following this, the next step involves merging the two features to utilize them collectively. Provided Data Tokenization is a technique in NLP that is employed to transform sequences into smaller units referred to as tokens. These codes vary in length, ranging from words to individual letters. The significance of this procedure lies in its ability to streamline the comprehension of human language by breaking it down into smaller, more easily analyzable components. Figure 1 illustrates the process of tokenization. Following this step, we employ pad sequences. The pad sequence's function is usually working in tasks that include sequences, such as NLP. To guarantee the equality of the input sequences, a fixed length of 512 Tokens was utilized in this study.

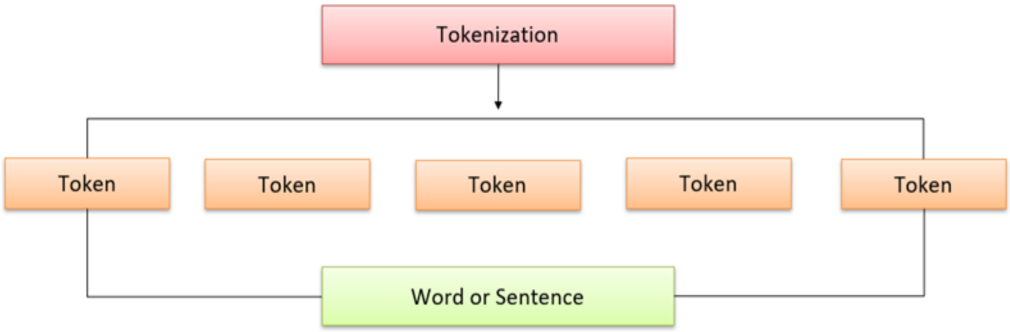

Figure 1. Pre-processing step in CNN-DNN model.

3.3. | Word Embeddings

It is essential to convert the input text into a numerical form, such as a vector or matrix, to input it into the network. Text can be represented as word vectors, with each word having its unique vector, known as word embeddings [10]. These word vectors are widely used for text representation.

Word embeddings are created through the process of training on a large corpus of text, usually focused on a specific language or domain. This allows the embeddings to capture the statistical correlations among all the words in the collection. Using publicly available pre-trained word embeddings is a more feasible strategy compared to training word embeddings from the beginning. Word2Vec and GloVe are the predominant pre-trained word embeddings, with Word2Vec being created by Google. The text is filtered by selectively retaining only the words that are present in the vocabulary while excluding those that are not.

3.4. | Convolutional Neural Network (CNN)

CNN utilizes matrix multiplication to produce outputs that are employed for subsequent training iterations. This method is known as convolution. This neural network is commonly known as CNN due to this specific rationale [11]. In the field of NLP, words in a sentence or news article are represented as word vectors. Afterward, these word vectors are used to train a CNN. The training is performed by explicitly defining the dimensions of the kernel and the number of filters. CNN possesses the ability to function in multiple dimensions [12].

Typically, a one-dimensional CNN is frequently used for tasks such as text classification or NLP. The Conv1D function is designed to process arrays that have a single dimension, particularly those that represent word vectors. Within a CNN, a filter with a window of a predetermined size is used to process the training data sequentially. In each iteration, the input is multiplied by the filter weights, and the resulting output is stored in an output array. The output array represents a feature map or output filter of the data. This technique is employed to ascertain a characteristic from the provided training data. This process can be graphically illustrated.

3.5. | Proposed Model

The CNN-DNN model consists of multiple layers. The initial layer, called the embedding layer, takes the article bodies and headlines as input, and converts each word into a 512-dimensional vector. With 128 features, this layer will produce a 512 by 128 matrix. The resulting matrix will have weights obtained through matrix multiplication, creating a vector for each word. The CNN layer then uses these vectors to extract contextual features, followed by Maxpooling, which reduces the data by selecting the maximum value from the feature map. Dropout layers are used to prevent overfitting. Subsequently, the flattened layer converts a multidimensional tensor into a one-dimensional vector. This is followed by two connected layers, then another dropout layer, and finally the output layer. Figure 2 illustrates this process.

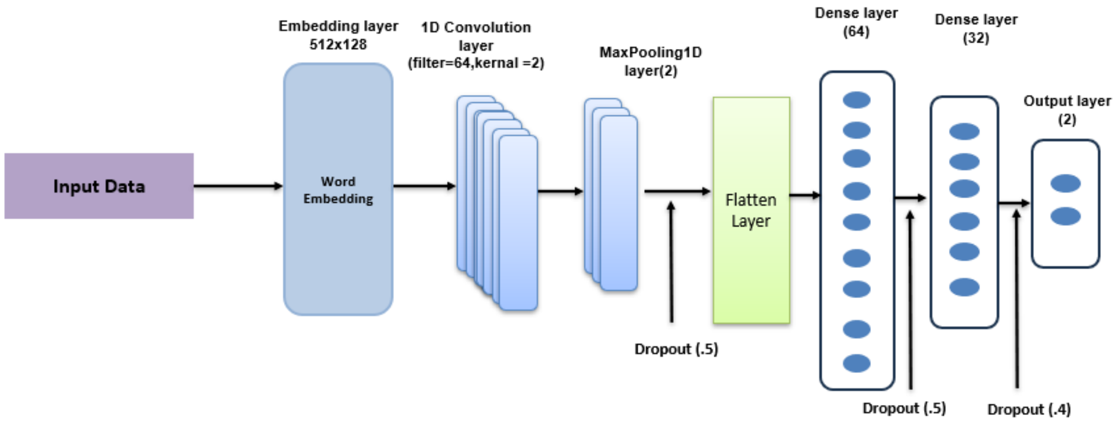

Figure 2. CNN-DNN model.

At this stage, DL models were created with the default parameter. All models were compiled to determine the used loss function. The Adam Optimizer was used to measure the error rate, and the evaluation metrics were used to evaluate its performance. The Categorical cross entropy (CCE) loss function is used to optimize the initial weights of certain DL models to increase classification accuracy. The loss function is mathematically defined as follows:

$$\text{Minimize: } \mathrm{loss(CCE)} = -\sum_{i=1}^{M} y_i \cdot \log \breve{y}_i \tag{1}$$

Where y_i is true value $\breve{y}_i$ is shorthand for a vector that contains all the outputs that were predicted based on the training samples.

3.6. | Training Deep Learning Models

The Adam optimizer with a learning rate of 0.0001 is utilized to train all DL models. The application of mini-batch gradient descent is employed to minimize the error, which is computed using the Categorical Cross-Entropy (CCE) loss function. By adjusting the weights based on a small subset of data from the training dataset. The training Data Set contains 43,000 news articles, and we utilize a batch size of 500. The mean weights change 86 times per epoch. After undergoing multiple training epochs using this approach, the DL models were then subjected to the evaluation stage, where their capacity to generalize is assessed. The key stages of training and evaluating the analyzed DL models for the Fake news Classification problem are illustrated in Figure 3.

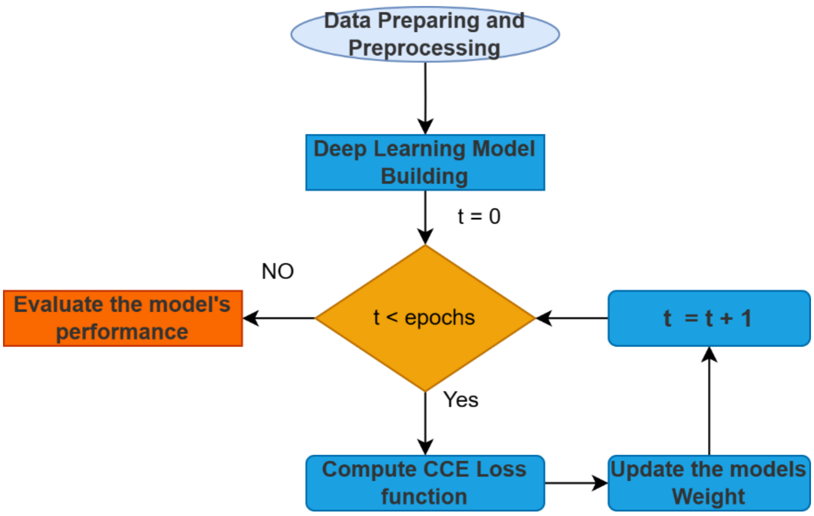

Figure 3. The general framework of training phase CNN-DNN model.

4. | Result and Discussion

The section presented extensively compares the performance of several DL models (LSTM model. Deep Model and Proposed Model) used in training the WELFake dataset.

4.1. | Evaluation metrics

The Comparisons are made through a set of metrics such as Accuracy, precision, recall, and F1-score.

- Accuracy: This metric is calculated from the number of correct predictions for all categories to the total number of predictions. The equation is as follows:

$$\mathrm{Acc} = \frac{(\mathrm{TPos} + \mathrm{TNeg})}{(\mathrm{TPos} + \mathrm{FPos} + \mathrm{TNeg} + \mathrm{FNeg})} \tag{2}$$

- Precision: This measure is calculated from the number of correct predictions for a category to the total number of predictions in the same category. The equation is as follows:

$$\mathrm{P} = \frac{TPos}{(TPos + FPos)} \tag{3}$$

- Recall: This statistic is used to display the proportion of accurately predicted samples for a class compared to all samples of the same class in a dataset. This metric can be calculated as follows.

$$\mathrm{R} = \frac{\mathrm{TPos}}{\mathrm{TPos} + \mathrm{FNeg}}. \tag{4}$$

- F1 Score: This is calculated by harmonic mean to balance between precision and recall. The equation is as follows:

$$F1\ \text{Score} = 2 \times \frac{\mathrm{R} \times \mathrm{P}}{\mathrm{R} + \mathrm{P}} \tag{5}$$

4.2. | Experimental Environment Setup

All Compared Models were implemented on the Kaggle environment with Nvidia Tesla P100 GPU and RAM of 16 GB, Python Version 3.9, and Keras Version 2.4. NumPy 1.19, and Pandas 1.2 are used for data operations. In addition, the Adam optimizer with a learning rate of 0.0001 and batch size of 500 is utilized for deep learning model training.

4.3. | Experimental Results

The experiments were conducted to evaluate the performance of our proposed deep-learning model for fake news detection. We compare its results with three models named LSTM, Gru, and Deep neural network models, Table 1 shows the performance of models with different metrics (accuracy, precision, recall, and F1-score), The suggested model achieved the best accuracy at 0.9732 the LSTM model achieved the lowest accuracy at 0.960. Figure 4 shows the rank of each model with different matrices. The Proposed model achieves the highest rank, followed by the Deep model. Figure 5 shows the Confusion Matrix used to describe the performance of a proposed model from a Visualizes and summarize it. Figure 6 shows the Accuracy and Loss Curve of the proposed model during the training process by evaluating each epoch on the validation dataset.

Table 1. Comparison between CNN-DNN and other DL models.

Model	Accuracy	Precision	Recall	F1 Score
LSTM	0.960	0.960	0.960	0.960
Gru	0.962	0.962	0.962	0.962
Deep	0.968	0.968	0.968	0.968
Proposed Model	0.973	0.973	0.973	0.973

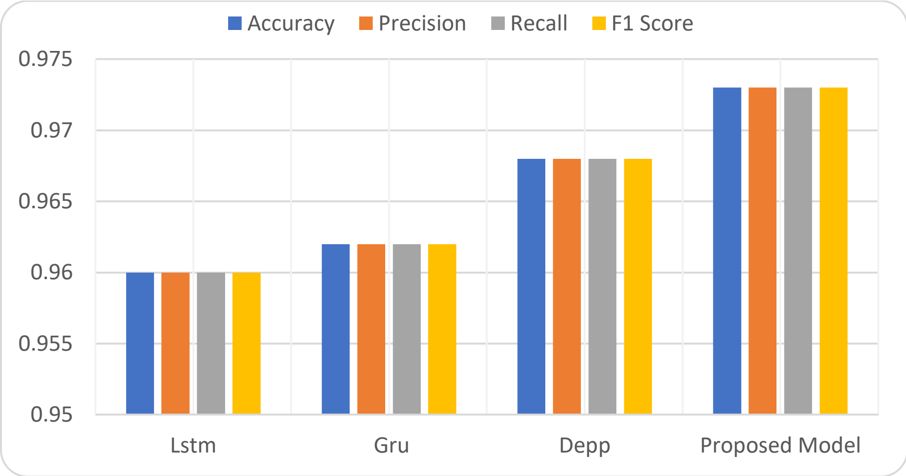

Figure 4. Shows the rank of each model.

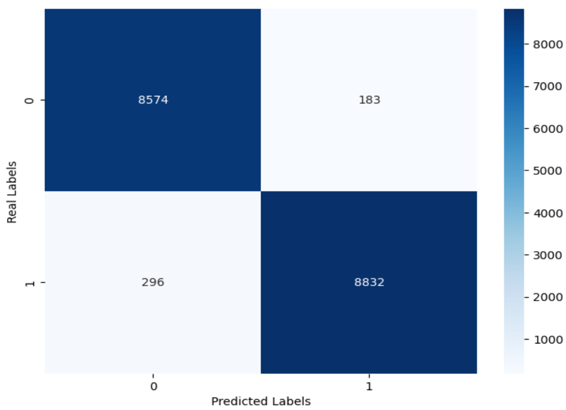

Figure 5. Confusion matrix of CNN-DNN.

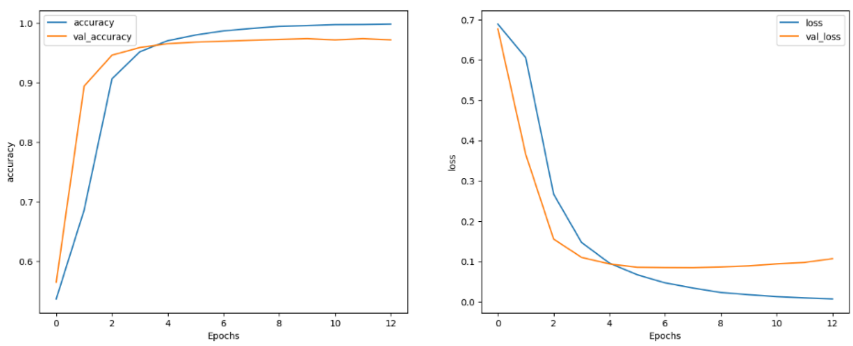

Figure 6. Accuracy and Loss Curve of CNN-DNN.

5. | Managerial Implications

To ensure the dissemination of accurate and reliable information, Egypt must successfully attain the objectives outlined in Vision 2030. In recent years, the field of FK detection has experienced notable progress, propelled by the progress made in NLP and ML. Despite significant progress, the task of detecting false information remains challenging and is an ongoing area of research, as spreading misinformation consistently alters their tactics. Nevertheless, technological progress offers promising methods to reduce the spread of misinformation and improve the reliability of online information. This study can be employed to optimize the process of precisely distinguishing counterfeit news. The proposed work can improve and facilitate the Media sector in realizing Egypt's vision for the year 2030.

6. | Conclusion and Future Work

This study introduced a model for detecting the stance of FKs, based on DL models and CNN. The proposed model comprises several layers. The first layer, known as the embedding layer, receives the article bodies and headlines as input and transforms each word into a vector with 512 dimensions. The CNN layer utilizes these vectors to extract contextual features, subsequently employing Maxpooling to reduce the data by selecting the maximum value from the feature map. Dropout layers are employed to mitigate the issue of overfitting. The Flatten layer transforms a tensor with multiple dimensions into a vector with a single dimension. Subsequently, there are two fully connected layers, succeeded by another dropout layer, and ultimately the output layer. This process yields promising results, achieving an accuracy score of up to 0.9732, which is significantly superior to the findings of previous studies.

Ultimately, we aim to expand our research by conducting a comparable examination on an entirely distinct dataset, such as Facebook. Through the process of categorizing false information disseminated on social media platforms, our objective is to make progress in the development of an automated system for detecting FKs.

Acknowledgments

The author is grateful to the editorial and reviewers, as well as the correspondent author, who offered assistance in the form of advice, assessment, and checking during the study period.

Funding

This research has no funding source.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy-preserving nature of the data but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Johnson, R.M., Social Media and Free Speech: A Collision Course That Threatens Democracy. Ohio Northern University Law Review, 2023. 49(2): p. 5.

[2] Altheneyan, A. and A. Alhadlaq, Big data ML-based fake news detection using distributed learning. IEEE Access, 2023. 11: p. 29447-29463.

[3] Palani, B., S. Elango, and V. Viswanathan K, CB-Fake: A multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT. Multimedia Tools and Applications, 2022. 81(4): p. 5587-5620.

[4] Sastrawan, I.K., I.P.A. Bayupati, and D.M.S. Arsa, Detection of fake news using deep learning CNN–RNN based methods. ICT express, 2022. 8(3): p. 396-408.

[5] Zhang, Q., et al., A deep learning-based fast fake news detection model for cyber-physical social services. Pattern Recognition Letters, 2023. 168: p. 31-38.

[6] Mohawesh, R., S. Maqsood, and Q. Althebyan, Multilingual deep learning framework for fake news detection using capsule neural network. Journal of Intelligent Information Systems, 2023. 60(3): p. 655-671.

[7] Soga, K., S. Yoshida, and M. Muneyasu, Exploiting stance similarity and graph neural networks for fake news detection. Pattern Recognition Letters, 2024. 177: p. 26-32.

[8] Abualigah, L., et al., Fake news detection using recurrent neural network based on bidirectional LSTM and GloVe. Social Network Analysis and Mining, 2024. 14(1): p. 1-16.

[9] dataset. Available from: https://www.kaggle.com/datasets/saurabhshahane/fake-news-classification/data.

[10] Li, Y. and T. Yang, Word embedding for understanding natural language: a survey. Guide to big data applications, 2018: p. 83-104.

[11] Zhang, X., X. Zhang, and W. Wang, Convolutional Neural Network, in Intelligent Information Processing with Matlab. 2023, Springer. p. 39-71.

[12] Nasir, J.A., O.S. Khan, and I. Varlamis, Fake news detection: A hybrid CNN-RNN based deep learning approach. International Journal of Information Management Data Insights, 2021. 1(1): p. 100007.