Al-Mustansiriyah Journal of Science
ISSN: 1814-635X (print), ISSN:2521-3520 (online) Volume 34, Issue 2, 2023 DOI: http://doi.org/10.23851/mjs.v34i2.1292

Review Article **Open Access**

Deep Learning for Fake News Detection: Literature Review

Mohammed Haqi Al-Tai[1], Bashar M. Nema[1*], Ali Al-Sherbaz[2]

[1]Department of Computer Science, College of Science, Mustansiriyah University, 10052 Baghdad, IRAQ.
[2]Cybersecurity & Computing Department, University of Gloucestershire, UK.

[*]Correspondent contact: bashar_sh77@uomustansiriyah.edu.iq

Article Info

Received
26/01/2023

Accepted
05/04/2023

Published
30/06/2023

ABSTRACT

The use of Deep Learning (DL) for identifying false or misleading information, known as fake news, is a growing area of research. Deep learning, a form of machine learning that utilizes algorithms to learn from large data sets, has shown promise in detecting fake news. The spread of fake news can cause significant harm to society economically, politically, and socially, and it has become increasingly important to find ways to detect and stop its spread. This paper examines current studies that use deep learning methods, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), as well as the multi-model approach, to detect fake news. It also looks at the use of word embedding models to convert text to vector representations and the datasets used for training models. Furthermore, the paper discusses the use of the attention mechanism in conjunction with deep learning to process sequential data.

KEYWORDS: Deep learning, fake news detection, natural language processing, machine learning, text classification, information credibility, social media analysis, semantic analysis, CNN, RNN, multimodal, datasets, word embedding, LSTM, Hybrid, BERT.

الخلاصة

يعد استخدام التعلم العميق لتحديد المعلومات الخاطئة أو المضللة ، والمعروفة باسم الأخبار المزيفة ، مجالًا متزايدًا للبحث. أظهر التعلم العميق ، وهو شكل من أشكال التعلم الآلي الذي يستخدم الخوارزميات للتعلم من مجموعات البيانات الكبيرة ، وعدًا في اكتشاف الأخبار المزيفة. يمكن أن يتسبب انتشار الأخبار المزيفة في إلحاق ضرر كبير بالمجتمع اقتصاديًا وسياسيًا واجتماعيًا ، وقد أصبح من المهم بشكل متزايد إيجاد طرق لاكتشاف انتشارها ووقف انتشارها. تبحث هذه الورقة البحثية في الدراسات الحالية التي تستخدم أساليب التعلم العميق ، مثل الشبكة العصبية التلافيفية (**CNN**) والشبكات العصبية المتكررة (**RNN**) ، بالإضافة إلى النهج متعدد النماذج ، للكشف عن الأخبار المزيفة. كما يبحث في استخدام نماذج تضمين الكلمات لتحويل النص إلى تمثيلات متجهية ومجموعات البيانات المستخدمة لنماذج التدريب. علاوة على ذلك ، تناقش الورقة استخدام آلية الانتباه بالتزامن مع التعلم العميق لمعالجة البيانات المتسلسلة.

INTRODUCTION

Social media has removed barriers to sharing and publishing information, providing users with limitless access to information. Consequently, social media has become a primary source of news and information, offering an effective means of gauging public opinion and tracking significant events like the COVID-19 pandemic. As a result, an increasing number of people are turning to social media platforms for information instead of traditional media outlets like newspapers and television. However, a considerable amount of news shared on social media is unverified and lacks reliable sources. Some individuals and organizations intentionally spread false information for their own gain. For instance, they may disseminate false information about the trading market to manipulate prices or use the rapid spread of fake news to advance their own beliefs [23]. Fake news is intentionally disseminated to mislead readers, often with skillful manipulation of language to play on readers' emotions and opinions [24]. The

negative impacts of fake news are well-documented and have far-reaching consequences. For instance, it can cause panic and misinformation during public health crises like the COVID-19 pandemic [10]. Additionally, fake news can influence public opinion and lead to social and political unrest, as seen in the 2016 U.S. presidential election [6]. Fake news has far-reaching negative impacts on various aspects of human life, making it one of the most significant research fields in artificial intelligence.

The proliferation of fake news on online platforms has become a significant challenge in recent years. To prevent the spread of fake news and its adverse effects, numerous strategies have been developed, including the use of deep learning (DL) methods, which have proven more effective than traditional machine learning methods in detecting fake news. DL is an emerging technology that has gained wide acceptance in research due to the availability of data programming frameworks such as TensorFlow and Keras [25]. Convolutional Neural Networks (CNNs) have achieved impressive results in classifying images and texts, while Recurrent Neural Networks (RNNs) have been instrumental in dealing with sequential data, such as text and speech. Furthermore, attention mechanisms have emerged as essential components of deep learning which has been increasingly employed in natural language processing (NLP) tasks like text summarization, and machine translation where it can assist the model in concentrating on pertinent words or phrases within a sentence.

The main objective of this paper is to inspect the latest developments, diverse architectures, and deep learning techniques that have been used to identify and prevent the spread of fake news.

DATASETS

To detect fake news using deep learning methods, a large, comprehensive, and reliable dataset that captures various aspects of fake news with a balanced distribution of both real and fake news articles is important. Accurate and reliable labeling of the dataset is essential for the deep learning model to learn effectively, and the dataset should have a wide range of features, such as linguistic and contextual features, that can be used to differentiate between real and fake news. This section examines some datasets that can be used to train and evaluate models for detecting fake news, including datasets with both short statements and full-length articles:

- FakeNewsNet: is a multi-dimensional data repository and constructer using a tool called FakeNewsTracker, available with a dataset for gathering, analyzing, and visualizing fake news detection on social media. FakeNewsNet consists of two data sets with content and social context, where the dataset contains 900 political news and 20K gossip news. The dataset is labeled only with two classes, fake or real [16].
- LIAR: this dataset consisted of 12.8k short human statements from fact-checking website politifact.com. The dataset labeled with six classes (pants-fire, false, barely-true, half-true, mostly-true, and true). In the LIAR dataset, the distribution of labels is relatively well balanced: except for 1,050 pants-fire cases, the instances for all other labels range from 2,063 to 2, 638. Dataset is partitioned into Training set size 10,269, Validation set size 1,284, Testing set size 1,283, Avg. statement length 17.9 [16, 26].
- PolitiFact: this dataset contains political news gathered from fact-checking website PolitiFact. The data contains more than 10k statements, where all statements are labeled into one of six classes (true, mostly true, half true, mostly false, false, and pants on fire); also, each statement are fact-checking by expert teams [26].
- GossipCop: GossipCop provides rated fact-checking entertainment stories gathered from entertainment websites [26].
- Pheme: a curated dataset consists of a collection of rumors and non-rumors posted on Twitter during breaking news, where the rumors labeled with their validity value, which is True, False, or Unverified [26,27].
- MediaEval: the dataset involves tweets associated with an event or place; it consists of 9k rumors and 6k non-rumor tweets gathered from different events. The data is a collection of images, videos, and text. If the image correlated to the event

corresponds to the tweet's text, then the tweet is labeled as “genuine”; otherwise, it is labeled as "Fake" [27].

- Weibo: this dataset is usually used with multi-model fake-news detection. Collected from verified rumor posts from May 2012 to January 2016 were from Weibo, a Chinese microblogging social network [27].
- BuzzFeed: a curated dataset represents a sample of news from 9 news agencies that were published on Facebook over a week close to the 2016 U.S. presidential election. Buzzfeed involves two datasets, one of fake news and another of real news; both have 91 observations and 12 features [27].
- SST (SST-2 and SST-5): The datasets were collected from movie reviews from rottentomatoes.com and consisted of 11,855 single sentences.SST-2 is labeled with two classes (positive or negative), while SST-5 labeled with five classes (very positive, positive, neutral, negative, and very negative). The datasets are usually used in the NLP community [27].

WORD EMBEDDING

- Word embeddings are a powerful tool in natural language processing (NLP) that allow words with similar meanings to be represented similarly. By mapping words from a higher dimensional to a lower dimensional space, word embeddings can capture the semantic relationships between words and improve the accuracy of predictions in various NLP tasks. There are several unsupervised techniques used for embedding input text into neural networks, including Word2Vec, Doc2Vec, FastText, and GloVe. [16,1]:
- Word2Vec: to learn word embedding using the Word2Vec method, two distinct learning models were presented (Bag-of-Words and Skip-Gram Model). In this technique, each word in the text is regarded as a context, and related words appear closer in the embedded lower-dimensional space.
- Doc2Vec: Doc2Vec is an extension of Word2Vec that learns vector representations of entire documents instead of just words. Tags the text and generates tag vectors where related words will have corresponding vectors nearest.
- FastText: each word is modeled by vectors, where each vector represents an n-gram.
- GloVe: assume an aggregate of word vectors associated with the probability of their co-occurrence in the corpus.

Word embedding has become an essential component of deep learning in NLP and has led to the development of language model-based embedding. Language models can train on large external datasets and represent a specific language model. One such model is Bidirectional Encoder Representations from Transformers (BERT), which has revolutionized NLP by delivering exceptional performance in a range of tasks, like sentence classification and text generation. BERT is a context-based model, which means that based on the meaning of the word in a sentence, it generates the embedding of the word, this is one of the strengths of BERT. Furthermore, BERT is based on the transformer architecture and can be understood as the transformer model. However, it differs in that it has only an encoder, while the transformer has both an encoder and decoder. The BERT model comes in two typical configurations, BERT-base with 12 encoder layers and BERT-large with 16 encoder layers; in both configurations, it can extract the embeddings from any BERT layers like last four hidden layers, last hidden layer, or from all hidden layers [28].

DEEP LEARNING METHODS

Before discussing Deep Learning methods in-depth, it is important to provide a brief overview of the concept of machine learning. Machine learning (ML) is a field of artificial intelligence that gives the capability of systems to solve problems by learning in a way that simulates human learning activities and acquiring new skills to improve performance. Moreover, Machine learning presents concepts that can solve complex problems in various tasks, like speech recognition, where the system has the ability to discriminate what we say. In the process of machine learning, data quality is the major factor in determining the efficiency

of the performance of machine learning algorithms [11]. Many researchers use different machine learning methods to detect fake news. By finding the most appropriate approach, many of them train models in good ways and achieve considerable accuracy in detecting fake news. In [2], the author proposes a model based on n-gram analysis and machine learning methods for detecting fake news. For creating features and representing the context of words, the author uses the N-gram Language Model word-based, also developing a classifier based on n-gram analysis to distinguish between fake and real news content; the idea is simply to use training data to create different sets of N-gram profiles to represent fake or real news content. The author used a dataset consisting of real news articles from Reuters.com and fake political news articles from kaggle.com and performed preprocessing on datasets which included (removing stop words, converting to lowercase, and punctuation removal) before representing by an N-gram model and vector representation. Furthermore, the author for features extraction uses two methods its Term Frequency (TF) and TF-IDF. After features extraction by using one of the two methods (TF or TF-IDF), it trains several machine learning algorithms on the dataset, where the dataset is split into 80% for training and 20% for testing. The author started the experiment by studying the effect of changing the size (n) of N-grams on the algorithm's performance, which starts with n=1 and increases the n value by one until n=4. Moreover, each value of n was tested with a different set of features. The proposed model achieved 98% accuracy. Furthermore, Linear-based classifiers and nonlinear with the same features extraction methods (TF or TF-IDF) and size of the N-grams (n=1 to 4), also by using top features ranging between 10000 to 50000, yielded considerable results, where Decision Trees achieved 89% accuracy and Linear SVM as 92% accuracy. Machine learning, despite its excellence in various tasks, however, performs poorly in some functions that seem easy for humans, such as distinguishing the voice if it is female or male. The poor performance of machine learning in some aspects led to the idea of simulating the work of neural networks of the human brain to adapt to learning and acquiring new knowledge. Now, the question is, what is deep learning? Deep learning has excelled at tasks where traditional machine learning approaches have performed poorly, and it has now become a widespread field for solving prediction problems beyond its original scope of computer vision and speech recognition [3]. Deep learning is making considerable progress in solving complex problems in natural language processing (NLP), it also introduces multilayered learning models by combining graphs with appropriate neural transformations. Many studies have used deep learning models to detect fake news [26]. This section reviews some deep-learning methods used to detect fake news.

1. Convolutional Neural Networks (CNNs)

A convolutional neural network is a kind of deep neural network that processes data with a grid-like pattern. It has an architecture inspired by biological data from the visual cortex. CNN is a mathematical construct comprising three types of layers: convolutional and pooling, for performing feature extraction, and a fully connected layer, for performing the final output. The convolutional layer consists of the convolution operation and activation function and represents the core of CNN that has a major role in the work of CNN. Moreover, the convolution layer contains filters that help to process a limited part of data at a time and is applied over all inputs. A pooling layer down sampling the dimensionality of the previous layer's feature maps (the feature map is the output of applying one filter to the preceding layer). The final layer in the architecture of CNN is a fully connected (FC) layer fed from the previous layers' output. It uses an activation function like (sigmoid or SoftMax) to produce the final outcome [7]. Convolutional neural networks (CNNs) are widely used in computer vision and have also been successful in natural language processing (NLP) tasks, especially in text classification. By arranging word vectors into a matrix and treating it as an image, CNNs can analyze groups of words together using a context window. The main advantage of CNNs is their ability to consider both individual words and sequences of contiguous words [22]. A lot of researchers use CNN architecture to detect fake news. In [17] the author proposes a deep CNN (FNDNet) model for fake news detection; it is composed of words embedded as an input

layer followed by a convolutional layer (conv1D) followed by a pooling layer (max-pooling) and finally fully connected layers. Instead of extracting hand-crafted features, the model is designed to extract features automatically through the deep neural network. The author uses a pre-trained word-embedding model called GloVe to generate vector representation used as input in the proposed model. GloVe is a widely used pre-trained word-embedding model that represents words as dense vectors in a high-dimensional space. Also, GloVe provides multiple pre-trained models with different sizes, each of which contains word vectors with a different number of dimensions. In this study, the author used the 'glove.6B.zip' model, which contains 822MB of data and represents each word as a 100-dimensional vector. The vectors generated from GloVe are fed to three parallel convolutional layers (conv1D) followed by a pooling layer (max-pooling layer). Additionally, the three parallel conv1D layers have different kernel sizes. The main reason for assigning different kernel sizes is to capture a greater amount of information during training, as different kernel sizes allow the model to capture different patterns in the input data. The pooling layer reduces the output from the previous layer, moreover it reduces the operations in the following layers. Therefore, in the proposed model, three max-pooling layers merge the output from the parallel convolutional layers (conv1D) then outputs from previous layers are concatenated and fed to the next convolutional layer. In the proposed model one flatten layer, the output generated from flatten layer is given to two dense layers to predict the final output. The model achieves considerable accuracy of 98.36% with the Kaggle news dataset.

In [12], the author proposed a deep two-path semi-supervised learning approach. The model contains three CNNs, one for supervised learning and the other for unsupervised learning, which are jointly trained to enhance detection performance using both labeled and unlabeled data. The author tested the model on the PHEME dataset, which includes 9 events, but only focused on 5 of them. Results showed that the proposed model achieved significant performance improvements compared to another baseline.

In [18] author proposed a multilevel-CNN blend of the local convolutional and global semantics features to capture semantic information from news articles to classify the news as fake or not. The author proposed a method for calculating the weight of sensitive words. The hierarchical CNNs make local semantic learning in convolutional layers possible, which helps capture sensitive words by adding consistency constraints in the semantics of global and local convolutional features. It encourages the model through regional semantics to focus more on sensitive words to obtain high classification accuracy. By testing the model on two datasets, the experiment yielded an accuracy of 91.67% on the Weibo dataset, and the experiment yielded an accuracy of the Liar dataset of 92.08%.

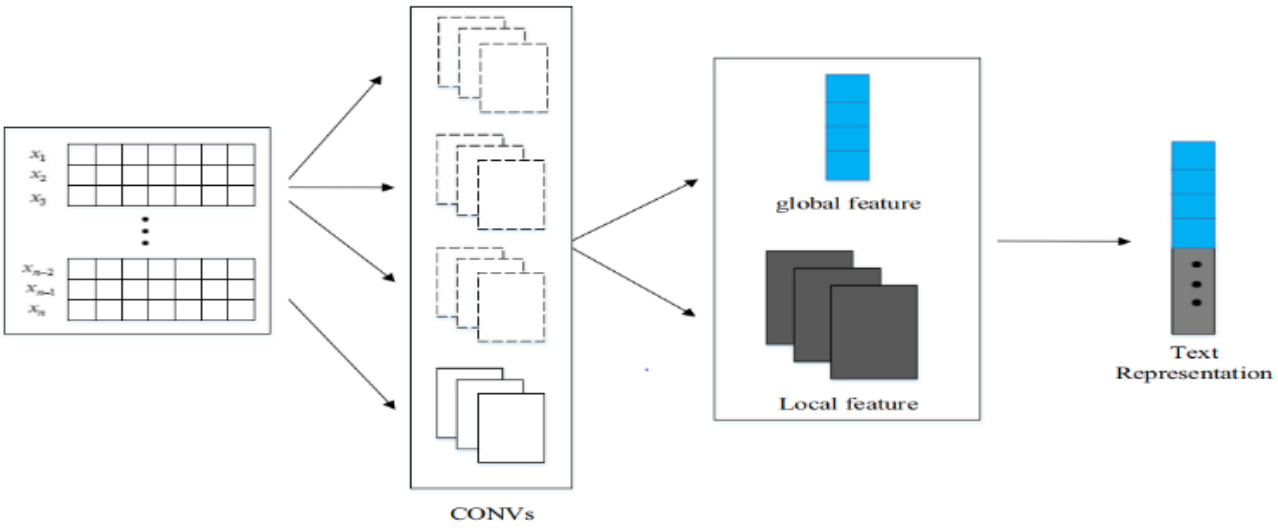

Figure 1. The proposed CNN architecture [18].

2. Recurrent Neural Networks (RNNs)

In a CNN, data flows without a feedback loop, which means it is not capable of handling sequential data like sentences, where the following values rely on the recent past values. For example, predicting the next word in a sentence depends heavily on the context of previous words. A neural network should be able to remember what happened moments before to deal with sequential data. RNN is a class of neural networks that deals with sequential data like text, speech, and time series, which have an arbitrary length. An RNN can be regarded as a neural network with memory that has the ability to remember states that happened moments before. The RNN architecture is composed of repeated cells in the time dimension, and each cell represents a single layer applied at each time step. Therefore, the depth of the network is determined by the length of the sequences or time steps, which also makes RNN contain a variable number of layers. Like other deep architectures, RNN is affected by the backpropagation of gradients. If an RNN deals with long sequences, the gradients computed during training can either vanish or explode during backpropagation through time (BPTT), causing slow data learning. In RNN architectures, the vanishing gradient problem has been relieved using Long Short-Term Memory (LSTM), which allows deep recurrent networks to train long sequences. The architecture of LSTM consists of multiple connected LSTM cells, each with a set of gates used to control data flow. The RNN has another cell known as the Gated recurrent unit (GRU), which is similar to LSTM but has an internal structure much simpler than LSTM, making it train faster. It also has two gates instead of three, such as in the LSTM cell [8, 13]. Both Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) give the deep neural network the capability to remember the previous state of the unit, especially in NLP. This property is helpful in processing sequences of sentences. Many researchers have utilized the flexibility of RNN in dealing with sequential patterns to detect fake news.

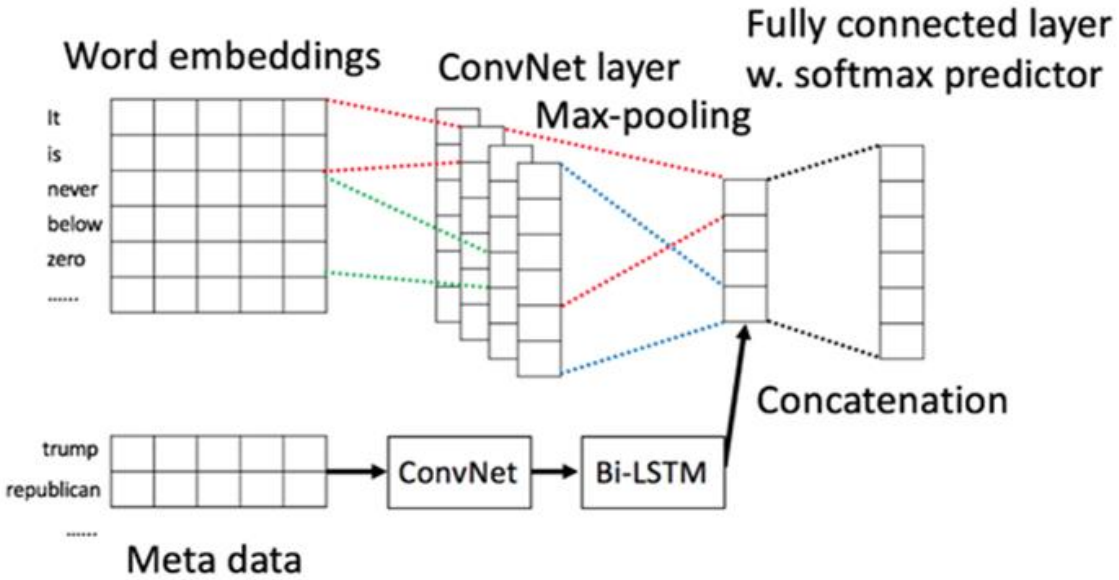

Figure 2. The proposed hybrid model [4].

In [4], the author presents the Liar dataset for fake news detection, consisting of 12.8K short statements manually labeled into six classes (pants-fire, false, barely-true, half-true, mostly-true, and true) collected from POLITIFACT.COM. The proposed neural network model is based on CNN and utilizes a randomly initialized matrix of embedding vectors to encode metadata embeddings. A convolutional layer is used to capture the relationship between text and metadata, followed by a max-pooling operation and a bi-directional LSTM layer. The max-pooled text vector is then blended with a metadata vector from the bidirectional LSTM and fed to a fully

connected layer. The final prediction is generated by the SoftMax activation function.

The author evaluated the model with various baselines using the Liar dataset. The majority baseline achieved about 0.204 accuracies on the validation and 0.208 accuracies on test sets. Standard text classifiers like regularized logistic regression classifier (LR) and support vector machine classifier (SVM) obtained significant accuracy. However, due to overfitting, the Bi-LSTMs performed poorly. The CNNs outperformed all models with an accuracy of 0.270. Thus, this work concludes that the combination of metadata with text achieves the best result in fake news detection.

In [19] author presents a hybrid model using CNN and RNN-LSTM for fake news detection. The dataset used with the proposed model consists of 216,682 news articles labeled with binary classification, where 1 if reliable news sources and 0 if un-reliable news sources. The author used a toolkit (NLTK toolkit) for preprocessing input text to remove special characters from the text; after this was done, CloVe word Embedding was used to map text to a vector representation. In the proposed model, CNN utilizes to extract high-level features from the text. Moreover, to enhance the performance of the proposed model, the author applied dropout technology with dense layers. Also, it utilized Two RNN-LSTMs to capture long-term and dependencies among word sequences as well as the context of the input sentence. To compare the efficiency of the hybrid model, the author ran the dataset on the CNN model, then on RNN-LSTMs separately, and finally on the hybrid model. In a result, the hybrid model achieved an accuracy of 92%. In contrast, both CNN and RNN achieved an accuracy of 90%.

3. Multimodal approach

The spread of digital transformation, such as social media platforms and websites, has contributed to the transformation of the press from publishing news on paper to multimedia news content, including video clips, images, and texts. This multimedia has immense importance associated with the credibility of the news and can exploit to detect fake news. Several studies utilize deep learning methods to learn hidden representations from text, images, video, and social context to detect fake news. In [9] author proposed a unified model named TI-CNN (Text and Image information-based CNN) to recognize fake news by analyzing text and image information using CNN. The dataset used in [9] consists of 20,015 news articles, divided into 11,941 fake news and 8,074 real news is used to build and test the model. Two parallel CNNs are applied to extract latent features from textual and visual content in the TI-CNN model. And then, to form a new set of representations of texts and images, both explicit and latent features are projected into the same space. Finally, a combination of textual and visual representation is used for news classification. Compared with several baseline methods, the performance of the proposed model outperforms significantly.

In [20] the author presents the Similarity-Aware FakE (SAFE) framework for detecting fake news by utilizing the relationship between textual and visual content. The method combines multimodal data to represent news articles and identify their falsity based on their text, images, or the mismatch between them. The model has three modules: the extraction of multimodal features, the modal-independent representation of textual and visual data, and the cross-modal similarity extraction to detect falsity by assessing the relevance of textual and visual data. The author used two datasets from PolitiFact and GossipCop, evaluated the performance of SAFE framework compared to significant baseline methods, and found that SAFE gives better result over both datasets. The proposed framework uses Text-CNN to extract textual features from news articles and pre-trained image2sentence models to process visual data within news content. By extending Text-CNN with a fully connected layer, the textual features are automatically extracted. The proposed model has the potential to improve fake news detection and prevent the spread of misinformation.

In [14], the author introduces SpotFake, a multimodal framework designed to identify instances of fake news by exploiting textual and visual information. SpotFake incorporates the

contextual information of textual data by utilizing BERT, a powerful language model that applies self-attention mechanism in each of its 12 encoding layers, to learn textual features. BERT is used alongside pre-trained VGG-19, which extracts visual features, and the combination of both features is used to create a representation vector for the news, allowing for accurate classification of fake news. The author uses Twitter and Weibo datasets to train the SpotFake framework. Performance of SpotFake outperforms compared to both EANN and MVAE on Twitter and Weibo datasets by 3.27% and 6.83%.

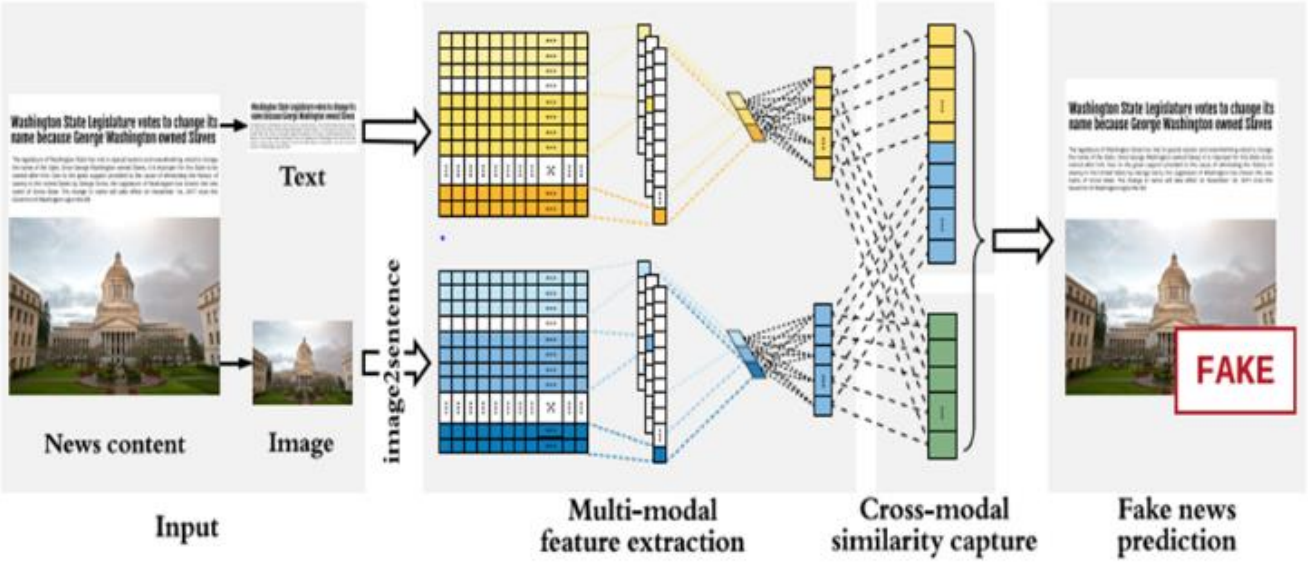

Figure 3. SAFE framework [20].

4. Attention Mechanism

From the above, the RNN, LSTM, and GRU are promising approaches for sequential modeling, such as language modeling. In research paper 'Attention Is All You Need' published by the Google Research and google brain team presents the concept of the attention mechanism used with recurrent networks, which in turn became an integral part of sequential modeling. The 'attention is all need' paper proposed a transformer model relying on an attention mechanism were composed of a stack of 6 layers. The model encoder comprises of 6 layers, each with two sub-layers: multi-head self-attention and a fully connected point-wise feedforward network. At the same time, a decoder is composed of 6 layers, each with three sub-layers (two attention sub-layer, fully connected point-wise feedforward network sub-layer). Moreover, the transformer model structure does not contain recurrent networks; in other words, recurrent has been abandoned and replaced with attention which requires increasing both the number of operations and also increasing space between two words. To accelerate the calculations a transformer model runs eight attention mechanisms in parallel with each attention sub-layer [5, 30].

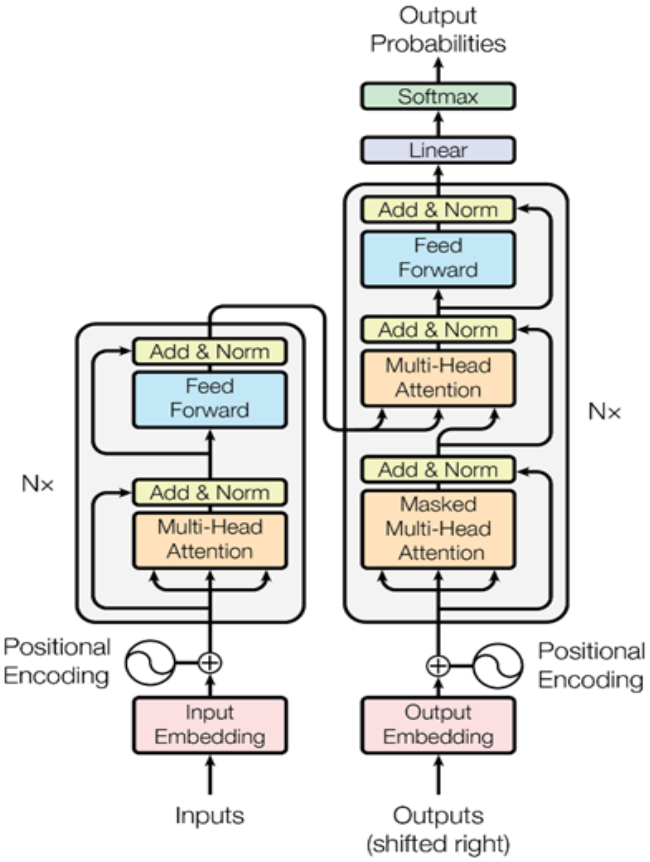

Figure 4. The Transformer - model architecture [5].

The attention mechanism addresses the problem of fixed-length of the encoder, it allows the decoder to access all hidden states of an encoder, and fixes limited access such as in traditional encoder-decoder. Also, the attention mechanism attempts to refine prediction by paying attention to various parts of the inputs. Furthermore, attention mechanisms can be soft or global-attention, local or hard-attention, and self-attention; the difference between these categories is how alignment is computed. For example, in self-Attention, alignment is computed on various sections of the sequence. Attention is a significant component in the "Embed, Encode, Predict" for building DL models for NLP [13].

In [31] author presents Attention-based C-BiLSTM model for detect fake news. The author uses the Liar dataset, which consists of 12.8k short statements labeled manual into six classes. Dataset pre-processes using several steps, like converting text to lowercase letters, removing special symbols, and tokenization. Then the output of pre-processing data is fed to the word embedding layer and, using the “GloVe” pre-trained word embedding model, generates a word vector. The author uses CNN to extract high-level features where the convolution layer applies the activation function RELU to process input data. After feature extraction via the Convolution layer is passed to a pooling layer. The pooling layer summarizes the features extracted in the input and constructs a new feature map. The author uses a maximum pooling layer and, with the help of filter size, determines the maximum value in a feature map. Moreover, a BiLSTM is introduced to learn the input sequence in both forward and backward. The attention mechanism layer is built on top of BiLSTM architecture for updating weights. In general, this mechanism's benefit is maintaining the longer input sequences. The attention layer is followed by Dense Layer, Flatten Layer, and SoftMax Layer respectively. The dense layer is providing all feature combinations from one layer to the next. The output of the flatten layer is applied to the output of the SoftMax layer. The SoftMax layer produces the final output. Using the same LIAR dataset, the proposed AC-BiLSTM model performs better than various state-of-art models.

In [15], the author introduces a hybrid model with a Self-attention mechanism for detect fake news. The proposed model can automatically capture contextual data's dependencies and learn global representation from contextual data for fake news detection. The model consists of two modules one for extracting the linguistic features and another for capturing the contextual. To extract the linguistic features from statements, the author used Text-CNN. Additionally, the author employed a self-attention mechanism to capture contextual information. This mechanism disregards the position of each component within the sequence and instead focuses on the overall representation of the sequence. The output from both the above modules are then combined and used as the representation of the news vector, which is then input into a fully connected layer for news classification. The author uses the Liar dataset to evaluate the performance of the model. The liar dataset contains 12.8k short statements labeled manual into six classes. Compared to LSTM-Attention and Hybrid-CNN, the proposed model achieved an accuracy of 45.3%, while both the LSTM-Attention and Hybrid-CNN models achieved an accuracy of 41.5% and 27.4%, respectively.

In [21] the author introduced a fake news detection model based on attention mechanism, which utilizes transformer model that leverages self-attention to improve efficiency, and also employs a typical encoder-decoder design. The encoder layer comprises of a self-attention layer and a feed-forward layer. The decoder has multiple layers; in addition to the self-attention layer and feed-forward layer, it contains an encoder-decoder attention layer it helps focus on relevant text elements. The transformer model needs to pay attention to both sentence and origin for Fake news detection. The final prediction generates via the SoftMax function. The author used Liar dataset to train the proposed model. The transformer model generally contains encoders and decoders. Each Encoder layer comprises a self-attention layer and a feed-forward layer. The decoder layer comprises both the self-attention layer and

feed-forward layer with an additional masked multi-head attention sublayer that includes the input from the encoder layer. compared to the hybrid-CNN baselines from (William YangWang's work in 2017), the proposed model achieved a 15% improvement in accuracy.

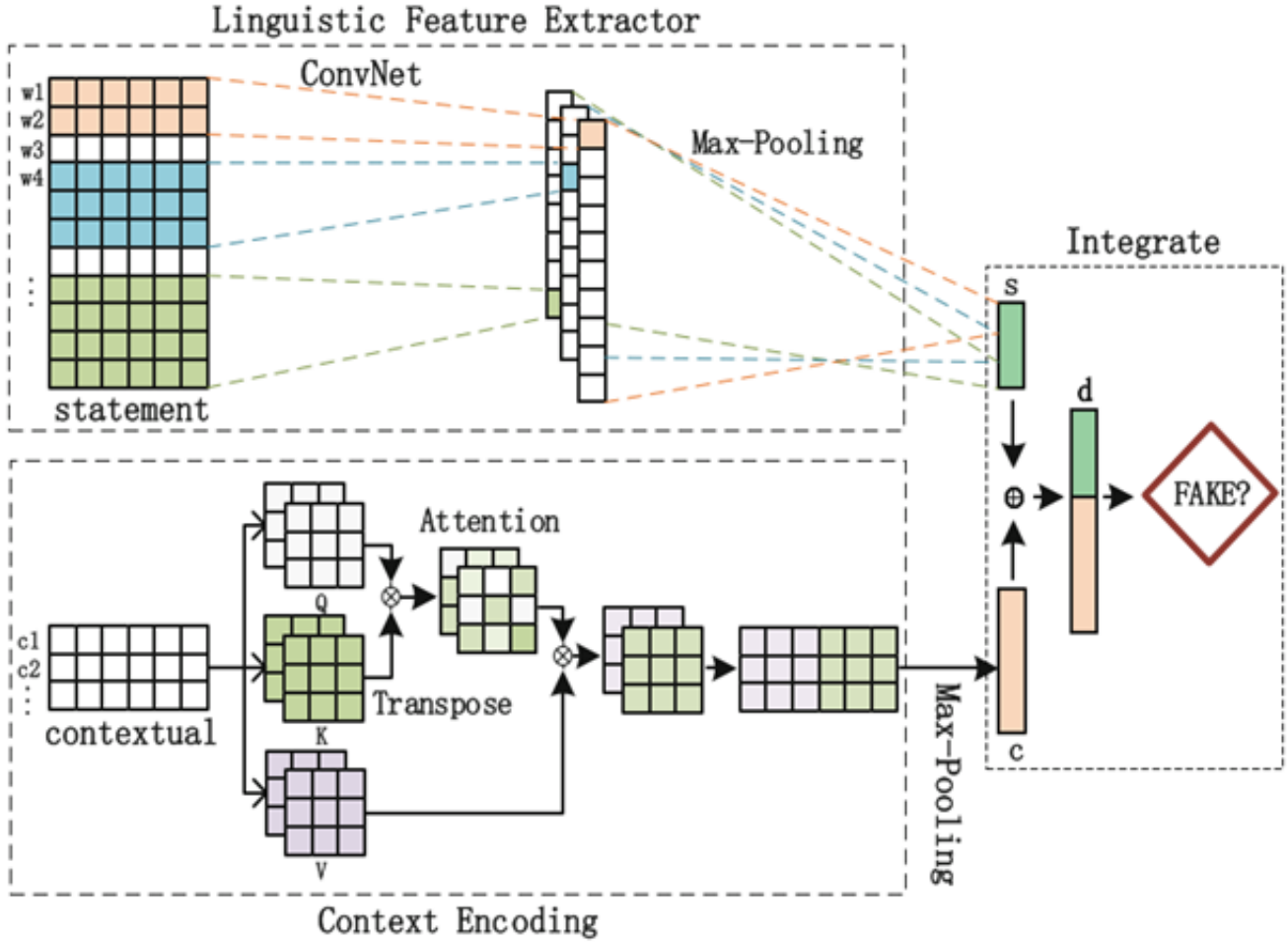

Figure 5. The structure of CMS [15].

Table 1. Summary of studies on detecting fake news.

Study/year	Deep learning model	Dataset	Accuracy precision/recall
[17] Kaliyar, Rohit. 2020	FNDNet: CNN	Kaggle news	Accuracy :98.36%
[12] Dong et al. 2019	Two-path CNN	PHEME	Recall: 77.58%
[18] Hu, Q.,2020	Multilevel-CNN	Weibo, Liar	Accuracy: 91.67% on the Weibo. Accuracy: 92.08% on the Liar.
[19] M. D. P. P. 2020	CNN + RNN-LSTM	SherLock-FakeNewsNet	Accuracy: 92%
[9] Yang, Y.,2018	TI-CNN: Text and Image based CNN	fake-news	Precision: 0.9220 Recall :0.9277 F1 score: 0.9210
[20] Zhou, X.,2020	SAFE: CNN+pretrained image2sentence model	PolitiFact, Gossip Cop	Accuracy:0.87% on the PolitiFact. Accuracy: 0.838 %. on the Gossip Cop
[31] Trueman, T.E.,2021	Attention-based CNN-BiLSTM	LIAR	Yielded Accuracy: (35.1%) and F1-score: (39%). More than other existing models
[15] Wang, Y.,2019	CMS: Text-CNN + Multi-head Self-attention	LIAR	CMS: 45.3% accuracy LSTM-Attention: 41.5% accuracy Hybrid-CNN: 27.4% accuracy
[21] M. Qazi,2020	attention-based transformer	LIAR	Accuracy: 0.4055%. Model improved fake news detection accuracy by 15% compared to the Hybrid CNN model.

CONCLUSIONS

We conclude that fake news has massive impacts on society as institutions and individuals. Many people fall victim to exploitation to unintentionally promote misleading news and direct their opinions to serve the interests of an institution or a cause. To reduce the risks and impact of this misinformation, many researchers have introduced concepts based on one of the essential branches of artificial intelligence, which is deep learning. Deep learning has revolutionized evolution in the fields of NLP. Researchers have presented promising hybrid models to reduce the spread of fake news, like CNN-RNN, and also multi-models based on exploiting the nature of fake news that the news can include text and images. At the same time, the word's meaning can change according to its position in the sentence and the effect of the preceding or following words. Thus, the model must be trained to deal with the context of the sentence. In this situation, the researchers exploited the concept of the attention mechanism to capture the relationship between words. The concept of the attention mechanism is one of the critical components of deep learning, especially in NLP. In the end, we conclude that deep learning concepts presented promising methods to reduce the impact of misleading information, as it achieved impressive results in this field.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to the Department of Computer Science at Mustansiriyah University for their invaluable support during the research and writing of this manuscript. We would also like to thank our colleague, Dr. Ali Al-Sherbaz from the University of Gloucestershire, UK, for his invaluable guidance and support throughout this project. Without his expertise and encouragement, this work would not have been possible. We are also grateful to the participants who volunteered their time and effort to make this research possible.

Disclosure and Conflict of Interest: The authors declare that they have no conflicts of interest.

REFERENCES

[1] Jason Brownlee. 2017. Deep Learning for Natural Language Processing. Machine Learning Mastery.

[2] Ahmed, H., Traore, I., Saad, S. (2017). Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. In: Traore, I., Lounging, I., Awad, A. (eds) Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments. ISDDC 2017. Lecture Notes in Computer Science (), vol 10618. Springer, Cham. https://doi.org/10.1007/978-3-319-69155-8_9

[3] Goodfellow, I., Bengio, Y., & Courville, A. (2017). Deep learning (adaptive computation and machine learning series). Cambridge Massachusetts, 321-359.

[4] Wang, William. (2017). "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. 422-426. 10.18653/v1/P17-2067. https://doi.org/10.18653/v1/P17-2067

[5] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., &Polosukhin, I. (2017). Attention Is All You Need. arXiv. https://doi.org/10.48550/arXiv.1706.03762.

[6] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of economic perspectives, 31(2), 211-236. https://doi.org/10.1257/jep.31.2.211

[7] Yamashita, R., Nishio, M., Do, R.K.G. et al. Convolutional neural networks: an overview and application in radiology. Insights Imaging 9, 611-629 (2018). https://doi.org/10.1007/s13244-018-0639-9

[8] Wei, Bhardwaj, A., & Wei, J. (2018). Deep Learning Essentials (1st ed.). Packt Publishing. Retrieved from https://www.perlego.com/book/578845/deep-learning-essentials-pdf (Original work published 2018)

[9] Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., & Yu, P. S. (2018). TI-CNN: Convolutional Neural Networks for Fake News Detection. arXiv. https://doi.org/10.48550/arXiv.1806.00749.

[10] Bode, L., & Vraga, E. K. (2018). See something, say something: Correction of global health misinformation on social media. Health communication, 33(9), 1131-1140. https://doi.org/10.1080/10410236.2017.1331312

[11] Moolayil, J. (2019). An Introduction to Deep Learning and Keras. In: Learn Keras for Deep Neural Networks. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-4240-7_1

[12] Dong, X., Victor, U., Chowdhury, S., & Qian, L. (2019). Deep Two-path Semi-supervised Learning for Fake News Detection. ArXiv, abs/1906.05659.

[13] Kapoor, Amita&Guili, Antonio & Pal, Sujit. (2019). Deep Learning with TensorFlow 2 and Keras: Regression, ConvNets, GANs, RNNs, NLP, and more with TensorFlow 2 and the Keras API, 2nd Edition.

[14] S. Singhal, R. R. Shah, T. Chakraborty, P. Kumaraguru and S. Satoh, "SpotFake: A Multi-modal Framework for Fake News Detection," 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), 2019, pp. 39-47. https://doi.org/10.1109/BigMM.2019.00-44

[15] Wang, Y., Han, H., Ding, Y., Wang, X., Liao, Q. (2019). Learning Contextual Features with Multi-head Self-attention for Fake News Detection. In: Xu, R.,

Wang, J., Zhang, LJ. (eds) Cognitive Computing - ICCC 2019. ICCC 2019. Lecture Notes in Computer Science (), vol 11518. Springer, Cham. https://doi.org/10.1007/978-3-030-23407-2_11

[16] Masciari, E., Moscato, V., Picariello, A., Sperli, G. (2020). A Deep Learning Approach to Fake News Detection. In: Helic, D., Leitner, G., Stettinger, M., Felfernig, A., Raś, Z.W. (eds) Foundations of Intelligent Systems. ISMIS 2020. Lecture Notes in Computer Science (), vol 12117. Springer, Cham. https://doi.org/10.1007/978-3-030-59491-6_11

[17] Kaliyar, Rohit & Goswami, Anurag & Narang, Pratik & Sinha, Soumendu. (2020). FNDNet- A Deep Convolutional Neural Network for Fake News Detection. Cognitive Systems Research. 61. https://doi.org/10.1016/j.cogsys.2019.12.005

[18] Hu, Q., Li, Q., Lu, Y. et al. Multi-level word features based on CNN for fake news detection in cultural communication. Pers UbiquitComput 24, 259-272 (2020). https://doi.org/10.1007/s00779-019-01289-y

[19] M. D. P. P. Goonathilake and P. P. N. V. Kumara, "CNN, RNN-LSTM Based Hybrid Approach to Detect State-of-the-Art Stance-Based Fake News on Social Media," 2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer), 2020, pp. 23-28. https://doi.org/10.1109/ICTer51097.2020.9325477

[20] Zhou, X., Wu, J., &Zafarani, R. (2020). SAFE: Similarity-Aware Multi-Modal Fake News Detection. arXiv. https://doi.org/10.1007/978-3-030-47436-2_27

[21] M. Qazi, M. U. S. Khan and M. Ali, "Detection of Fake News Using Transformer Model," 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2020, pp. 1-6. https://doi.org/10.1109/iCoMET48670.2020.9074071

[22] Vajjala, S., Majumder, B., Gupta, A., & Surana, H. (2020). Practical natural language processing: a comprehensive guide to building real-world NLP systems. O'Reilly Media.

[23] Albahar, Marwan. (2021). A hybrid model for fake news detection: Leveraging news content and user comments in fake news. IET Information Security. 15. https://doi.org/10.1049/ise2.12021

[24] Petratos, Pythagoras. (2021). Misinformation, disinformation, and fake news: Cyber risks to business. Business Horizons. 64. https://doi.org/10.1016/j.bushor.2021.07.012

[25] Mridha, M. F., Keya, A. J., Hamid, M. A., Monowar, M. M., & Rahman, M. S. (2021). A Comprehensive Review on Fake News Detection with Deep Learning. IEEE Access. https://doi.org/10.1109/ACCESS.2021.3129329

[26] G, S.K. (2021). Deep Learning for Fake News Detection. In: Data Science for Fake News. The Information Retrieval Series, vol 42. Springer, Cham. https://doi.org/10.1007/978-3-030-62696-9_4

[27] Chakraborty, T. (2021). Multi-modal Fake News Detection. In: Data Science for Fake News. The Information Retrieval Series, vol 42. Springer, Cham. https://doi.org/10.1007/978-3-030-62696-9_3

[28] Ravichandiran, S. (2021). Getting Started with Google BERT: Build and train state-of-the-art natural language processing models using BERT. Packt Publishing Ltd.

[29] G, S.K. (2021). Deep Learning for Fake News Detection. In: Data Science for Fake News. The Information Retrieval Series, vol 42. Springer, Cham. https://doi.org/10.1007/978-3-030-62696-9_4

[30] Rothman, D. (2021). Transformers for Natural Language Processing: Build innovative deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and more. Packt Publishing Ltd.

[31] Trueman, T.E., J, A., Narayanasamy, P., & Vidya, J. (2021). Attention-based C-BiLSTM for fake news detection. Appl. Soft Comput., 110, 107600. https://doi.org/10.1016/j.asoc.2021.107600

How to Cite

M. . Haqi Al-Tai, B. M. Nema, and A. . Al-Sherbaz, “Deep Learning for Fake News Detection: Literature Review”, *Al-Mustansiriyah Journal of Science*, vol. 34, no. 2, pp. 70–81, Jun. 2023.