Fake News Detection with a Novel Long Short-Term Memory and Bahdanau Attention Mechanism

David An
AP Research
Greater Latrobe Senior High School
davidzhongtaian@gmail.com

Keywords: Machine Learning, AI, Natural Language Processing

Abstract

The rapid development of social media and online news outlets has accelerated the spread of fake news across the internet. The accessibility and convenience of social media has further driven the drastic change of information consumption. As a consequence, fake news has become a significant concern because of 1) its inevitable exposure to large populations and 2) the potential to cause significant damage in modern society. Fake news most commonly utilizes extreme examples to catch a reader's attention [1]. The current trend is utilizing deep leaning techniques and pipeline methods to create an effective way to discriminate fake news. In the paper, we propose a model named CNN-BAM (Convolutional Neural Network with Bahdanau Attention Mechanism network) that can detect between fake and authentic news utilizing CNNs and an attention mechanism. The proposed approach is staged in three parts. The first stage is a data preprocessing stage. The second is adapting a convolutional neural network architecture for the problem. The last stage is testing the model. In order to validate and benchmark our results, we performed a series of metric tests on the MIT Fake News Dataset (mit.fakenews.edu) and achieved a precision rate of 90.31% as well as an accuracy rate of 88.56%. Our experiments indicate the model is more effective at fake news detection. However, this model still has a significant amount of room to improve in future research.

Introduction

In recent years, the accessibility of news and media consumption has changed drastically [1]. Along with this development, different journalistic organizations looking to generate traffic to their websites also have drastically changed how their information is presented [1]. The convenience of web and social media allows people to quickly know what is happening and absorb vast quantities of information [2], [3]. With over 71% of American adolescents, ages 13-17, using Facebook, the explosion of social media accounts in the past decade has only accelerated the speed at which information spreads [2], [4]. Social media not only expands people's global view of the world but their view of their friends [4]. An abundance of articles on the web means that a small amount may be replaced with an "alternative narrative" [5]. With the large audience base and near-instant speeds of dissemination on social media, fake news and misinformation campaigns can quickly go viral [2], [4].

In this "social media explosion," information management and checking has become a highly tedious task [2]. Constantly improving tools and technologies prevents traditional technologies from working efficiently in fake news detection (FND) [5]. Now, misinformation can spread farther, quicker, and deeper within groups in a phenomenon known as the Echo Chamber Effect [2], [5]. Users on social platforms tend to follow other like-minded people on social networks and thus form a network of people with similar narratives [2], [6]. Facebook users tend to follow other like-minded users which in turn "echoes" narratives within a network [6], [7]. When an increasing amount of people accept fake news and believe the content to be true, fake news can affect financial, political, science, and education sectors [2]. For example, after the 2020 election, misinformation campaigns that were formed about the outcome dominated social and mainstream media months after the result. In addition to that, the rumors fueled by online conspiracy theorists such as PizzaGate or the Clinton Administration "Pedophile Sex Ring" led to real life consequences that posed a threat to society [8].

In the explosive world of social media and digital news outlets, fake news detection still remains a challenging task. Firstly, many models developed tend to side and train with more biased outlets which may not transfer to effective results on mainstream media [3], [9]. In addition to that, fake news is becoming ‘smarter’ as it is no longer ‘black and white’ but instead acting as a ‘double-edged sword’ that mixes a variety of true and false information [2], [9], [10]. For example, media outlets intending to release fake news may cite true statistics which greatly increases the difficulty of modeling and prediction based on the content [2].

Mainstream media such as CNN and New York Times (NYT) have had varying degrees of success when combating fake news and misinformation. One such method was attaching a confidence index for each article or columnar article published [9], [11]. Machine learning methods such as Shallow-Wide CNN and Ensemble Learning have also achieved significant results in detection and recognition. Convolutional neural networks were originally used in image processing and recognition but only recently have they been used in text processing and learning [12]. GloVe (global vector representations) embeddings also help build a more accurate classifier as it utilizes highly trained embeddings for turning text representation into vector representations. This feature allows a network to grab features from raw data [2].

In this work, the effectiveness of a model that utilizes a CNN (convolutional neural network) and GloVe embeddings was studied. To ensure accuracy or results, a public dataset was used and the results were compared to other public benchmarks [2]. To test the effectiveness in a real-life setting, the model will be tested against unseen testing examples that are unrelated to training examples. Finally, a pipeline is proposed to automate the detection of fake news. The paper contributions are summarized below:

- A machine learning model is built which can be generalized to a spectrum of news, not just satirical content.
- The proposed model takes advantage of convolutional filters along with word embeddings to detect fake news.
- The proposed model achieved significant results compared to previous generation models. In addition to that, the model performed much faster (in terms of run time)

Related Work

Fake news detection has become a recent issue that has become extremely studied and attracted significant attention already [13]. Although the concept of fake news has been in existence since the rise of mass media and press, the meaning of it has changed significantly [1]. Some authors describe “fake” having the connotation of being “deceptive” or “misleading” while other authors believe it means an “outright lie” [1], [5]. In this study, we define fake news as detected articles that contain misleading and/or falsified information [2], [12]. Also, there have been many previous approaches to fake news detection with methods such as Long Short-Term Memory (LSTM), Multi-Head Attention Mechanisms (MHAM), and Bi-directional Language Models (BiLMs). However, they have lacked the ability to capture spatial ability to an extent that is accurate. At this point, this section will summarize previous relevant research on fake news detection and different models and how they could be improved upon in the process.

Content-based Fake News Detection

The current research in fake news detection has increased dramatically in recent years as an attempt to curb the spread of misinformation. A majority of this work has been concentrated solely on the medium it spreads on: social media and web applications [14]. The current nature of technology allows fake news to spread easily: someone sees a fake news article, forwards it to their friend, and so on [15]. Content based fake news detection is also one of the simpler forms of text classification. Content based fake news detection is in the name. It detects based on the contents , or text of the article, This method is also known as fact checking [16].

In this study, we will mainly focus on content-based fake news detection: extracting features from the content of the article. Usually, the content comprises of the article title and body text. Content based fake news detection is looking at the content of the website or article and solely basing it off the content. This is different from the field of context based fake news detection where followers, time, location, and international events are taken into consideration.

Long Short-Term Memory

Long Short-Term Memory (LSTM) was initially applied to the field of text processing. These systems are an extension of Recurrent Neural Networks (RNNs) and address the downfalls of RNNs. With its built-in mechanisms, this extension of RNNs can effectively store representations in cell-like structures in order to capture the overarching pattern of a body of text [17]. Because of its characteristics, many researchers have applied this method to text classification with significant results [9], [17]. Now, the LSTM structure is widely applied on a variety of natural language processing tasks.

Every news article has a “main idea,” which is the topic of the text and is reinforced throughout the article [1]. LSTM networks have excelled in capturing the general idea of an article with a memory mechanism. For news articles, the main idea is sometimes not present in the first sentence, but instead, throughout the whole article. Therefore, it is suitable to use a mechanism that captures the idea over a period of time. For fake news detection, the content of the article is used. Previously, text classification often used RNNs and LSTMs and other deep learning methods for content extraction and detection [2], [17]. Although these networks have produced significant results, they still have some downfalls. For example, the basic LSTM structure does address vanishing gradients. however, they fail to remove it completely as data still moves in a linear manner. In addition to that, LSTMs require large amounts of resources to train which does not make it a feasible candidate for study with current resources [18]. Therefore, we try to shy away from using LSTMs in our experiment in order to improve model precision and efficiency.

Previous Models

Convolutional Neural Networks and SMHA-CNN

Convolutional Neural Network (CNN) is an architecture in computer vision that has become dominant in a variety of domains. While it was initially designed for image processing, it is now applied across a variety of tasks [19]. CNNs usually have three main building blocks: convolutional, pooling, and fully connected layers. A typical CNN will usually have several of these layers stacked on each other [19].

Self Multi-Head Attention Convolutional Neural Networks are proposed with the attention mechanism to capture content across a span of text. Like CNNs, the attention mechanism was initially applied to the field of image processing because of its ability to capture local contextual features. Because of its characteristics, Bahdanau applied the attention mechanism to natural language processing (NLP) and produced significant results [2], [20]. However, the basic attention mechanism fails to draw relationships between individual words. Vaswani et al. proposed the self multi-head attention mechanism which learns the mathematical relationship between words and alleviates the current situation [2], [21].

Bidirectional Long Short-Term Memory Model

As stated above, Long Short-Term Memory models (LSTM) are structures that learn the state of the model before outputting data [11]. While this is an extension of the traditional RNN, it presents many shortcomings. One of the most significant ones is that a RNN is only able to make use of content that was previously mentioned in close proximity to the current cell [11]. To solve this problem, a bidirectional RNN model is proposed where two layers are stacked parallel to each other and receive subsequent information [11], [22]. With a forward and backward running RNN, it allows the model to receive information from both directions [11]. By applying the same concepts to LSTM models produces a Bidirectional LSTM (BiLSTM) model which can capture context over a wider area [2], [11], [22]. While this does provide a greater range of data, gaps are also formed in the spatial range of the model as the proximity is still close to one another [11].

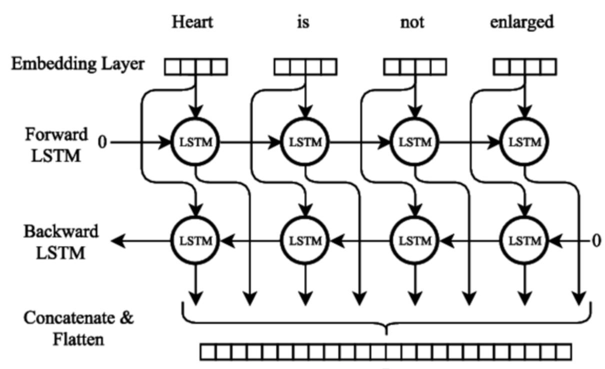

Fig. 1. Bidirectional-LSTM model. Adapted from Cornegruta et al., "Modelling radiological language with bidirectional long short-term memory networks," *Proceedings of the Seventh International Workshop on Health Text Mining and Information Analysis,* pp 17-27 [23].

The Proposed Model Architecture

In the literature review, it seemed that much of the problem focused on the lack of ability to close the spatial range gap in text classification. [24] stated that the design of certain models made for text classification have one flaw: the lack of spatial range. In addition to that, models such as LSTMs also incorporate vanishing gradients and other errors which can lead to inaccurate results. Originally, the attention mechanism was fitted for image processing where the need of spatial attention across pixels was greatest. However, we propose that the attention mechanism can be adapted to the text classification field as well with certain models. We anticipate that the attention mechanism can address the problem created by LSTMs and other networks mentioned above.

To detect fake news based on the style and content, a neural network model architecture of a long short-term memory with an attention mechanism is proposed. Specifically, we use the Bahdanau Attention Mechanism proposed in [20]. We hypothesize that if we combine the spatial aspect of an attention mechanism with text processing, we can increase the efficiency and accuracy of the task. The Bahdanau Attention Mechanism is perfectly suited for taking a weighted sum of hidden states and generating contextual representations for words [25]. This fits the problem perfectly as being able to generate a contextual representation for a news article will prove to be immensely useful. Although it was originally designed for image classification, the features and concepts can be adapted to the text as well. Combined with a long short-term memory module, this can be valuable to remember relationships between words. With this hypothesis, we hope to gain positive results from our experimentation.

Word Embeddings

Word embeddings are widely used in natural language processing. By mapping a word into a real valued vector, one can measure the semantic difference and observe the spatial relationship between words. By using a pre trained or self-developed model, we can word embed effectively. Word embedding is shown in Eq. 1. Doing so for each word creates a matrix $X \in R^{d\,x\,n}$ with any news article.

$$X = (x_1, \ x_2, \cdots x_n)$$

Where $x_n \in R^d$ where there are d dimensions word vectors corresponding to the $\boldsymbol{n}$-th vector in the news [2].

The main word embeddings used are known as GLoVe (Global Vector Representations) [26]. GLoVe is an unsupervised learning algorithm that was specially designed for obtaining vector representations of words [24]. By using pretrained word corpus as well as vector spaces, it allows us to better train our model as well as provide better results overall. Studies such as Yang et al. also vectorize the data before using the data to train the model.

Bahdanau Attention Filter

Word embeddings generate a word matrix and feature representation for the word. One downside is that the vector representation only is the definition of that certain vector and cannot generate relationships between relational vectors [24]. The attention mechanism (also known as Bahdanau Attention Filter) proposed by [20] can determine the spatial relationship between words and effectively generate better results [2], [19]. This attention filter can effectively evaluate sentences in neural machine translation but this mechanism will be adapted for the study's needs [20].

The Bahdanau Attention Filter can be described as the following: given a source sequence **x** with a length of *n*, we output our targeted classification of **y** of length *m*.

$$x = [x_1, x_2, \cdots, x_n]$$
$$y = [y_1, y_2, \cdots, y_n]$$

Experiments

The experiment was done in the style of an experimental research methodology. During the study, variables were manipulated, and quantitative data was collected which fits the concept of experimental research. Besides quantitative data, qualitative data was collected as well to triangulate the results. The qualitative data that was observed was the news publishers and context of the news setting. Being able to use both types of data allows us to draw an informed opinion on our model as well.

Dataset and environment

The experiment uses 20,000 articles scraped by [27] as the dataset. The data was present in two files (separated by [\n]). According to the publishers of the dataset: "The fake news dataset consisted of approximately 12,000 articles pulled from Kaggle, which maintains a blacklist of fake news source websites. Our real news dataset consisted of 9,000 Guardian articles and over 2,000 New York Times articles [27]." In addition to that, this Kaggle dataset is commonly found in many other different papers as well such as [4], [6], and [12]. Since we are studying the characteristics of fake news and how we can develop a model to detect it, it is appropriate to use a tailored dataset designed by professionals in the field to help accelerate the learning and development process.

After preprocessing, the experiment used a dataset of 19,525 x 2 to train and evaluate the model. The experimental environment is shown in Table 1. The use of Keras and other libraries is based on convenience and past knowledge in the area.

Items	**Configuration**
OS	macOS 11.2 Big Sur Beta
System Configuration	CPU: Intel i7-5650U, RAM: 8G
	GPU: Intel HD Graphics 6000, 1538 MB
Python libraries used	Keras, nltk, numpy, pandas, Scikit-learn,

Table 1: Environment configuration

Problem Definition

In a formal manner, the problem is defined as following: given a news article *A,* the objective is to generate a function *f* such as:

$$f(a) = \begin{cases} 0 & \text{if } a \text{ is true} \\ 1 & \text{if } a \text{ is false} \end{cases}$$

Since there are only two outcomes in the experiment, this can be defined as a binary classification problem. In addition to that Rodriguez and his researchers used binary classification in a similar model citing the "convenience" it provided during the evaluation process [14].

Metrics

The news articles are classified into either real or fake, making this a binary classification problem. This means there can be only two distinct categories that the news can fall into. During the experiment, a variety of metrics were used to observe accuracy and performance of the designed model. Metrics are an industry standard to use in order to evaluate machine learning models. To make the experiment as close to the true industry as possible, we will also be using metrics shown in the following equation [2], [6], [10]:

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F - 1\,Score = \frac{2*Precision*Recall}{(Precision+Recall)}$$

Using these metrics, we can calculate values in order to compare our model to industry benchmarks. By doing so allows us to see how well it performs to other models and lets us see the effect of the inclusion of an attention mechanism.

Experiments, Results, Discussions

Data Preprocessing

The training data was presented in the form of raw text. Because it is raw text, there was a large quantity of useless text that does nothing to help contribute to the accuracy of the model. Many examples of this useless text are the following: image captions, hyperlinks, referral links, advertisement text, and headers and footers. Much of the data processing utilized various packages in BeautifulSoup. BeautifulSoup is a freely downloaded python package specifically designed for text preprocessing and is widely used in the field of natural language processing.

First, REGEX was used to clean punctuation, metadata, and advertisement outlines. The text is then read into a .CSV format file while utilizing the error_bad_lines argument to segment incomplete or corrupted files. Second, the text was iterated and cleaned with the Natural Language Tool Kit (NLTK) stopwords file. By using NLTK, it effectively filters out words that do not contribute to the semantic meaning of the text such as "he", "she", "and", etc. [28]. Finally, to create equal processing, the vector length was calculated. The average article was around 244 words. Since the average is the most representative of the whole sample, that was the value we used as the max length for training. This is done as statistically, the average is representative of the sample length as a whole and has the best representation as well. All of these methods are available in the packages above.

Word Embedding

During the experiment, the Keras *Tokenizer()* was utilized as word embeddings. Instead of using a pretrained word embedding file, *Tokenizer()* created an unique vector for each news entry. By doing so, excess vectors aren't created and loaded into memory; therefore, saving computing power. In addition to that, word embedding allows text to be turned into "machine readable" format. Embedding with tokenizers is commonly done in other machine learning papers as well which justifies our method of embedding the articles [5-7]. We feed in each news entry as a vector of integers which make up the news article.

Model Design

To train the model, a 2-1 split was employed. This meant that 66.67% of the data was used to train the model while 33% of the data was used to test and validate the model. The LSTM model was designed with 12 units and an input shape of 64. In addition to this, to prevent overfitting, a dropout value of 0.25 was used alongside the Rectified Learning Unit (ReLU) activation function. Finally, a dense layer was added to combine all the outputs. The model was compiled with the binary cross-entropy loss function and ADAM optimizer. In total, the model had over 700,000 decision nodes. It was trained for a total of 10 epochs and in batch sizes of 80. These metrics were chosen as they seemed to be a common starting point for many studies [2], [17]. In addition to that, these starting points also follow the standard machine learning pipeline in the literature as well. A majority of researchers in the field collect and clean data before designing and evaluating it—a standard procedure.

Evaluation and Results

In this experiment, the result is shown in Table 2: Results. It was shown that the Bahadanau Attention Mechanism with an LSTM achieved a final and maximum precision of 90.71% as well as an accuracy value of 88.56%. In our test and validation set, the model performed well with high accuracy rates. To see how it compares to previous models, several baseline results were found and compared against. Apart from classic metrics, we compare our model to variations of models as well such as the BiDirectional Long Short Term Memory network or a Gated Recurrent Unit Network. During the experimentation process, the results were measured by several metrics, namely accuracy and precision. These values were calculated with built in methods and packages during the coding process.

	Precision	**Accuracy**
Bi-LSTM	86.20%	86.10%
LSTM-1200	67.30%	67.50%
GRU	82.30%	81.20%
Experimental	**90.31%**	**88.56%**

Figure 2: Results of experimentation

Through experimentation, we learned that the inclusion of an attention mechanism was indeed effective in increasing text classification efficiency. This is shown with significant gains across precision and accuracy. The results will be discussed further in the next section.

Discussions

In the project, the proposed method utilizing the Bidirectional Long Short-Term Memory in combination with the Bahdanau Attention Mechanism achieved more efficient results compared to industry standards. This is compared to the Bi-LSTM, LSTM 1200, and GRU. During experimentation, the proposed method reached a 90.71% accuracy. This value is significantly higher than other benchmarks such as the ones stated above (Figure 2). However, during the experimentation process, there were a few limitations that could have hindered a better outcome. Since cloud computing was not accessible during the experimentation process, all the calculations were performed on a MacBook Air 2017. Without the specialized hardware equipment such as accelerated processing units and graphics processing units, there were not enough computational resources and power to train and develop more complex models. If more computing power were accessible, the results may have been even more significant. In addition to that, a more complex model could have paved the way for better performance metrics. Also, a more complex model could have meant less error. This will generate better performance for the model as well.

Computational power wasn't the only limitation in the experiment. The processing size of the dataset was also another limiting factor. Because of the size of texts, filler words such as "and," "or," and other contextual words were filtered out. This was done to ensure better processing and to redirect attention onto the content of the article. Even by cutting out the filler words, the text remained significantly too large to process. In the end, the article was limited to 140 words. Even with this limitation, it still covered most of the articles as 89% of news segments were 140 words or less [6].

From this study, it was shown that attention mechanisms are effective ways for capturing the meaning of a series across time. By capturing spatial attention, the study showed that a deep learning system has the ability to learn the meaning of a series. In addition to that, the system now isn't limited to just adjacent tokens but can have spanned "attention" as well.

Future Applications

This deep learning model augmented with the Bahdanau Attention Mechanism achieved baseline scores that performed better than many baseline models. In the results, the proposed model performed 1.3x better than some models in addition to being faster overall. This supports our initial hypothesis that the inclusion of the attention mechanism does increase text classification efficiency. However, even with a 90.71% accuracy as well as an 88.56% precision rate, we still have not reached a level of significance that will prove to be effective in the real world. This indicates that the model has a lot of room available for possible improvement. Future research into attention mechanisms can investigate how to improve that result.

The detection and removal of fake news on social media platforms has become an extremely challenging task now. In this paper, we investigate a solution to this problem. Throughout our study, we offer a deep learning system-based model and evaluate the effectiveness of it. In addition to that, we use two standard performance metrics to compare our model against industry benchmarks. The system outperforms the LSTM-1200, GRU, as well as other systems. This suggests that the attention mechanism is indeed useful for increasing accuracy, supporting our hypothesis. Our research also allows for more insight into using attention for deep learning. All in all, it is shown that the inclusion of the attention mechanism is indeed useful in increasing text classification accuracy.

To clean the increasingly busy environment of the internet, the need for a fake news detection system is urgent. In the paper, we propose an effective model that utilizes the Bahdanau attention-based Convolutional Neural Network. When detecting a new topic, the new method performed extremely well. Also, we can see that the filler words were not essential to text classification or towards the main "idea" of the text. In addition to that, we compared the experimental result with benchmark results and observed that it performed significantly better.

In future work, a possible research direction can be experimenting with further applications into the natural language processing sector and see whether the attention mechanism can be used to process series more efficiently. It is suggested that capturing spatial attention can be useful to increase efficiency. In addition to that, in the future, we would like to increase the depth of the model by incorporating an increased number of features such as semantic and syntaxical detection. Also, researchers can possibly include outside knowledge (general information) into their models as well [29]. In an industrial standpoint, the system designed here can be potentially used in online media platforms where posting isn't limited. In the paper, it showed the effectiveness of the attention mechanism in a series-like data. We also hypothesize that future research into other time series such as market forecasting as well as weather forecasting with the attention mechanism can yield positive results as well. This way, fake news can be scanned and reviewed effectively prior to take down. In addition to that, a continuous pipeline can be built for the machine to constantly improve itself and become more accurate.

We would like to thank Mrs. Kimberly Mamaril for her supervision and guidance throughout the whole academic year on this project.

References

[1] J. P. Baptista and A. Gradim, "Understanding fake news consumption: A review," *Soc. Sci.*, vol. 9, no. 10, pp. 1–22, 2020, doi: 10.3390/socsci9100185.

[2] Y. Fang, J. Gao, C. Huang, H. Peng, and R. Wu, "Self Multi-Head Attention-based Convolutional Neural Networks for fake news detection," *PLoS One*, vol. 14, no. 9, pp. 1–13, 2019, doi: 10.1371/journal.pone.0222713.

[3] H. Ahmed, I. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification," *Secur. Priv.*, vol. 1, no. 1, p. e9, 2018, doi:

10.1002/spy2.9.
[4] S. P. Christensen, “Social media use and its impact on relationships and emotions,” *All Theses Diss.*, pp. 1–73, 2018, [Online]. Available: https://scholarsarchive.byu.edu/etd/6927.
[5] J. Albright, “Welcome to the Era of Fake News,” *Media Commun.*, vol. 5, no. 2, pp. 87–89, 2017, doi: 10.17645/mac.v5i2.977.
[6] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake news detection on social media: A data mining perspective,” *arXiv*, no. i, 2017, doi: 10.1145/3137597.3137600.
[7] M. Del Vicario *et al.*, “Echo Chambers: Emotional Contagion and Group Polarization on Facebook,” *Sci. Rep.*, vol. 6, pp. 1–13, 2016, doi: 10.1038/srep37825.
[8] Y. Tsfati, H. G. Boomgaarden, J. Strömbäck, R. Vliegenthart, A. Damstra, and E. Lindgren, “Causes and consequences of mainstream media dissemination of fake news: literature review and synthesis,” *Ann. Int. Commun. Assoc.*, vol. 44, no. 2, pp. 157–173, 2020, doi: 10.1080/23808985.2020.1759443.
[9] C. V. Meneses Silva, R. Silva Fontes, and M. Colaço Júnior, “Intelligent Fake News Detection: A Systematic Mapping,” *J. Appl. Secur. Res.*, vol. 0, no. 0, pp. 1–22, 2020, doi: 10.1080/19361610.2020.1761224.
[10] D. Vora and K. Iyer, “Evaluating the Effectiveness of Machine Learning Algorithms in Predictive Modelling,” *Int. J. Eng. Technol.*, vol. 7, no. 3.4, p. 197, 2018, doi: 10.14419/ijet.v7i3.4.16773.
[11] D.-H. Lee, Y.-R. Kim, S.-M. Park, and Y.-J. Yang, “Fake News Detection with Deep Learning,” *J. Inf. Process. Syst.*, vol. 15, no. 5, pp. 1119–1130, 2019, doi: 10.3745/JIPS.04.0142.
[12] M. Z. Amin and N. Nadeem, “Convolutional neural network: Text classification model for open domain question answering system,” *arXiv*, 2018.
[13] F. A. Ozbay and B. Alatas, “A novel approach for detection of fake news on social media using metaheuristic optimization algorithms,” *Elektron. ir Elektrotechnika*, vol. 25, no. 4, pp. 62–67, 2019, doi: 10.5755/j01.eie.25.4.23972.
[14] Á. I. Rodríguez and L. L. Iglesias, “Fake News Detection Using Deep Learning,” *arXiv*, 2019, doi: 10.35940/ijitee.i7059.079920.
[15] D. de Beer and M. Matthee, *Approaches to Identify Fake News: A Systematic Literature Review*, vol. 136, no. Macaulay 2018. Springer International Publishing, 2021.
[16] J. Z. Pan, S. Pavlova, C. Li, N. Li, Y. Li, and J. Liu, “Content based fake news detection using knowledge graphs,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11136 LNCS, pp. 669–683, 2018, doi: 10.1007/978-3-030-00671-6_39.
[17] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
[18] “Understanding of LSTM Networks,” *GeeksforGeeks*, 2020. https://www.geeksforgeeks.org/understanding-of-lstm-networks/ (accessed Jan. 14, 2021).
[19] A. Patil and M. Rane, “Convolutional Neural Networks: An Overview and Its Applications in Pattern Recognition,” *Smart Innov. Syst. Technol.*, vol. 195, pp. 21–30, 2021, doi: 10.1007/978-981-15-7078-0_3.
[20] D. Bahdanau, K. H. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.
[21] A. Vaswani *et al.*, “Attention is all you need,” *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.
[22] S. Martina, S. Gherardini, and F. Caruso, “Machine learning approach for quantum non-Markovian noise classification,” pp. 1–14, 2021, [Online]. Available: http://arxiv.org/abs/2101.03221.

[23] S. Cornegruta, R. Bakewell, S. Withey, and G. Montana, "Modelling Radiological Language with Bidirectional Long Short-Term Memory Networks," *Proc. Seventh Int. Work. Heal. Text Min. Inf. Anal.*, pp. 17–27, 2016, doi: 10.18653/v1/w16-6103.

[24] F. Yang, A. Mukherjee, and E. Gragut, "Satirical News Detection and Analysis using Attention Mechanism and Linguistic Features," *arXiv*, pp. 1979–1989, 2017.

[25] A. Express, "A Comprehensive Guide to Attention Mechanism in Deep Learning for Everyone," *Analytics*, 2019. https://www.analyticsvidhya.com/blog/2019/11/comprehensive-guide-attention-mechanism-deep-learning/.

[26] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global Vectors for Word Representation," *Stanford*, 2015. https://nlp.stanford.edu/projects/glove/.

[27] MIT, "Fake News Detector," 2017.

[28] "https://www.nltk.org." .

[29] T. Saikh, A. Anand, A. Ekbal, and P. Bhattacharyya, *A Novel Approach Towards Fake News Detection: Deep Learning Augmented with Textual Entailment Features*, vol. 11608 LNCS, no. June. Springer International Publishing, 2019.