Automatic Identification of Fake News Using Deep Learning

Ethar Qawasmeh, Mais Tawalbeh, Malak Abdullah
College of Computer and Information Technology
Jordan University of Science and Technology
Email: {eaqawasmeh16,matawalbeh18}@cit.just.edu.jo,mabdullah@just.edu.jo

***Abstract*—The rapid development of computing trends, wireless communications, and the smart devices industry has contributed to the widespread of the internet. People can access internet services and applications from anywhere in the world at any time. There is no doubt that these technological advances have made our lives easier and saved our time and efforts. On the other side, we should admit that there is a misuse of internet and its applications including online platforms. As an example, online platforms have been involved in spreading fake news all over the world to serve certain purposes (political, economic, or social media). Detecting fake news is considered one of the hard challenges in term of the existing content-based analysis of traditional methods. Recently, the performance of neural network models have outperformed traditional machine learning methods due to the outstanding ability of feature extraction. Still, there is a lack of research work on detecting fake news in news and time critical events. Therefore, in this paper, we have investigated the automatic identification of fake news over online communication platforms. Moreover, We propose an automatic identification of fake news using modern machine learning techniques. The proposed model is a bidirectional LSTM concatenated model that is applied on the FNC-1 dataset with 85.3% accuracy performance.**

***Index Terms*—fake news detection, social media, machine learning, deep learning, Fake news challenge**

I. Introduction

The recent trends and advances in communication and mobile technologies along with the widespread of the Internet have simplified accessing the news all over the world. Moreover, the uncontrolled development in the web-based life applications and the race of sharing and spreading news-related data between the international organizations in the field of social correspondence have affected the reliability of the news [1]. Recently, social media has become one of the main sources of information. The main factors beyond this are the low cost, the speed of access, ease of use, and availability on all digital devices including desktops, smartphones, iPod and others.

The "fake news" concept refers to intentionally disseminate false information on social media that aims to confuse and mislead the reader to achieve economic or political agendas. In addition, the diverse and growing number of industry players in the field of news writing and spreading have led to creating news articles that are difficult to know whether they are credible or not. [2].

The massive growth of fake news in social media has motivated researchers in academic institutions and industrial domains to obtain solutions that limit this phenomenon [3]. The widespread of fake news that preceded the United States presidential elections 2016 is considered a controversial issue that affected the public opinions [4]. The risk of catastrophic impacts of the rapid spread of false news over the social network sites is increasing dramatically. Therefore, spreading fake news is a worldwide annoying issue and many countries are criminalizing the creation and the distribution of misinformation online [1].

Automatically detecting fake news is considered a challenge for the existing content-based analysis methods. There is an urgent need for investigating machine learning approaches to detect fake news. Several fake news detection models and approaches have been used to identify fake news including traditional learning [5, 6] and deep learning models [7, 8]. There are three main categories for these approaches which are: content, social context, and propagation [9]. The existing neural network models have outperformed the traditional ones on their performance due to the outstanding ability of feature extraction, but still, these methods are not able to detect fake news, newly arose news, and time-critical events [10].

The remainder of this paper is organized as follows. Section 2 gives a brief description of existing works in detecting fake news. Section 3 provides a background about fake news definition and detection. Section 4 proposes a system architecture to determine the presence of fake in an article. Section 5 presents the evaluations and the results. Finally, section 6 concludes with future directions for this research.

II. Related Work

There are a significant number of studies has been conducted for fake news detection in the context of machine learning. Pérez-Rosas et al. [11] constructed two new datasets that covered seven news domains. One dataset was collected by crowd-sourcing, which covered six news data set. The second dataset was collected directly from the web and covered one domain. They conducted several exploratory analyses in order to identify linguistic features that were presented in fake news content and the differences from real news. They built a fake news detector using linear SVM classifier and five-fold cross-validation based on the combination of lexical, syntactic, and semantic information as linguistic features. Based on their results, their model achieved accuracy up to 78%.

[1] https://www.poynter.org/ifcn/anti-misinformation-actions/

978-1-7281-2946-4/19/$31.00 ©2019 IEEE

Davis and Proctor [12] proposed a model of bag-of-words followed by a three-layer multi-layer perceptron (BoW MLP). This model achieved the best results compared with the other three neural architecture models in their study. They used The Fake News Challenge dataset (FNC-1) which was presented in a public competition that aimed to find automatic methods for detecting fake news. The objective was to classify the dataset that consists of headline-text pairs as unrelated, agreeing, disagreeing, or discussing. They achieved 93% of accuracy. In addition, Miller and Oswalt [13] used the same dataset for detecting fake news using a neural network model with attention mechanism. They built a network architecture using multiple Bidirectional LSTMs and an attention mechanism to predict the entailment of the articles to their paired headlines. The best result was achieved by the BiLSTM + MLP (Multilayer Perceptron) with 57% accuracy. Other trials with attention models performed about 55% accurately.

Advances in artificial intelligence have made it easy to create fake visual contents, like images and videos, that are hard to be spotted if real or fake. These visual contents can be easily used to accelerate the spread of fake news through social media. In 2018, Wang et al. [14] proposed a framework named Event Adversarial Neural Network (EANN), which aimed to derive event-invariant features and can identify fake news based on multi-modal features and learn transferable feature. For evaluation of the performance for their model, they collected two real social data-sets from Twitter and Weibo. The experimental results showed that their proposed EANN model outperformed the state-of-the-art methods at that time.

In 2018, Zhang et al. [15] built a deep diffusive network model named FACE DETECTOR relied on a set of explicit and latent features extracted from the textual information. The proposed model aimed to detect the labels of news articles, creators and subjects simultaneously. Furthermore, they had performed an extensive experiment on a real-world fake news dataset from PolitiFact to compare FACE DETECTOR with several state-of-the-art models. They concluded that the proposed model had outstanding performance in identifying the fake news articles, creators, and subjects in the network.

Recently, in 2019, Monti et al. [16] proposed a fake news detection model based on geometric deep learning on a Twitter social network. They collected news articles from the archive of news that were provided from popular fact-checking organizations such as Snopes1, PolitiFact2, and Buzzfeed3. They filtered out all data which not-contained at least one URL reference on Twitter. Their experiments showed that the social network features, such as structure and propagation, achieved high accuracy (92.7 % ROC AUC) on fake news detection model.

III. Fake news

The broadcasting media of news has been changed drastically from newsprint, telephone, radio, and television to the internet by online news and social media. In this section, we discuss the main definitions of fake news. We also explore the fake news detection methods, especially stances detection, which has been used in this paper.

A. Problem definition

In the past, the news was being broadcast by speech between two or more parties. Then, the printed newspapers became wandering the world as trusted sources that follow strict codes of practice. Before the born of internet, we got our news from traditional sources which come from professionals in the form of newspapers, radio or television. However, the internet has enabled a whole new way to publish and broadcast information and news with less regulation or editorial standards. Three years ago, Fake news became a common concept that has grown in popularity. Fake news is defined as a news that intentionally includes false information and aims to mislead the reader [17]. This definition has two main characteristics: authenticity and intent. In the first characteristic, fake news includes false information varied in their content that can be verified. The second one, fake news is created with dishonest intention aims to mislead consumers. This definition of fake news has been widely adopted in recent researches [18, 19].

B. Stance Detection

The rapid growth of players number in news spreading field, especially on social media, has led to an increase in the number of news which are hard to tell if they are credible or not. The common way to highlight a specific type of news is by choosing a headline that grasps the attention of readers. Unfortunately, the news content does not always reflect the headline. The researchers studied this problem by analyzing several datasets that are focused on stance and headline detection task to identify whether a news headline is - in reality- related or unrelated to the corresponding news article. A further investigation had been applied on data to determine if the news headline is agreed, disagree or discussed with the content of the corresponding news article. The dataset used in this current paper was provided by Fake News Challenge (FNC-Stage 1) [2]. This challenge was prepared in 2016 by a category of academics and volunteers with a goal of addressing the fake news problem. FNC-1's experimental setup and the top three results are discussed by a group of researchers in [20]. The same data was studied and analyzed using different methods by different researchers [21, 22]. Also, the attention module was applied on the same data in 2017 by Chopra et al. [22] and they got 51% accuracy. In addition to FNC-1 datasets, other datasets were proposed for the same purpose of detecting fake news [23, 24].

IV. Approach

This section demonstrates the used dataset. In addition, we discuss the data cleaning process and the feature extraction method that are used in our approach. Then, we present the model architectures.

[2]Fake news challenge, http://www.fakenewschallenge.org/

A. Dataset

We have used a specific dataset that serves our goal to address the fake news issue, distinguishes untrustworthy news, and improves automatic detectors tools. The dataset contains news articles that are labeled with a class label to imply whether the article is trustworthy or not. The dataset is derived from the Emergent Dataset created by Craig Silverman [25] and is used in February 2017 for Fake News Challenge Stage 1 (FNC-1). The dataset can be downloaded from their Github page[3]. The dataset is provided as two CSV's files { train_bodies and train_stances }. Table I provides the dataset format.

TABLE I: Dataset format

FNC-1 Dataset	
train_bodies (1683 articles)	
Field	**Description**
Article Body	The body text of article
Body ID	The Article ID
train_stances (49972 headlines)	
Field	**Description**
Body ID	The Article ID
Headline	The article headline
Stance	The lable

Due to the differentiation in the number of articles and headlines in the dataset files, we have used many-to-many mapping. As a result, we got 49972 pairs of headline and body texts, each with a corresponding class label. Table II and Fig 1 show the description and distribution rate of stance classes in train _stances file. We assumed that all features in our used dataset are important, so we didn't apply cleaning process (removing stop words, etc..). We split the dataset into train, validation, and test. The test data was separated before training process and has been only used for final model evaluation. The distribution of splitting dataset as follows: 60%, 20%, 20% to training, validation, and test sets, respectively.

TABLE II: Stance classes description in train_stances file

Label	**Description**
Agree	The body text agrees with the headline.
Disagree	The body text disagrees with the headline.
Discuss	The body text discusses the same topic as the headline, but does not take a position
Unrelated	The body text discusses a different topic than the headline

B. Feature extraction

In order to extract features from text data, we have transformed the inputs into vector space using the pre-trained GoogleNews[4] word vector model with 300-dimension that are taken from 3 billion running words of Wikipedia.

Its worth mentioning that we also used the Gensim pre-trained word embedding model for word representation, which applied "dropping stop words" as a cleaning data preprocessing. Using word embedding on our final classifications module

[3]Fake news challenge data-set, https://github.com/FakeNewsChallenge/fnc-1

[4]word2vec-GoogleNews-vectors, https://github.com/mmihaltz/word2vec-GoogleNews-vectors

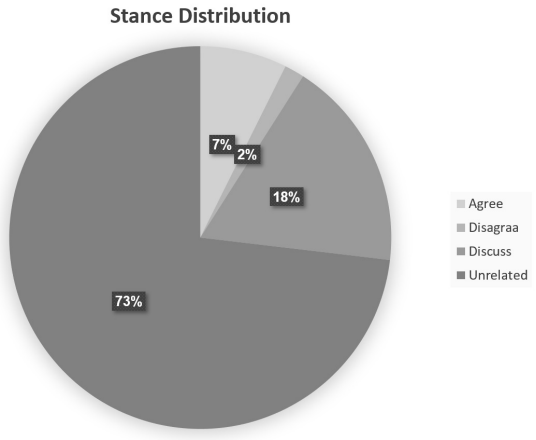


Fig. 1: Stance classes distribution rate in train_stances file

architectures, the Word2Vec approach with Google's pre-trained word embedding model achieved the best results. Table III illustrates a simple comparison between these approaches.

TABLE III: Pre-trained word embedding models

Model	**Description**	**Unique Words**	**Vector Length**
Gensim	Is trained in our data-set, which is roughly 49972 records	23977 words	(100-300) features
Google News	Is trained on roughly 3 billions of words from a Google News data-set	3 Billions words	300 features

The term **Word embedding** refers to mapping the words into vectors of real numbers as initial weights, then the improvement of the weight values occurres at the embedding layer, which needs a time. Using the **Word2Vec** allows us to use a pre-trained word embedding model. So, the improvement of the weight values happens before the embedding layer, which reduces the training time inside the final classification model[26].

C. Model architectures

After running over tens of experiments across different neural network architectures and hyper-parameters, two models shows satisfying results. These two models are FND_Bidirectional LSTM concatenated model and FND_Multihead LSTM model that are both described in details (FND refers to Fake News Detection).

FND_Bidirectional LSTM concatenated model: The headline and article texts are converted into two separated embedding layers using Google's pre-trained model. The result of the two embedding vectors are concatenated to feed the model as shown in Figure 2. The model consists of two CNN's layers with 32 and 64 filters, respectively. The CNNs are followed by max pooling to avoid over-fitting[5] before passing it through a Bidirectional LSTM layer with 100 memory units.

[5]pooling, http://ufldl.stanford.edu/tutorial/supervised/Pooling/

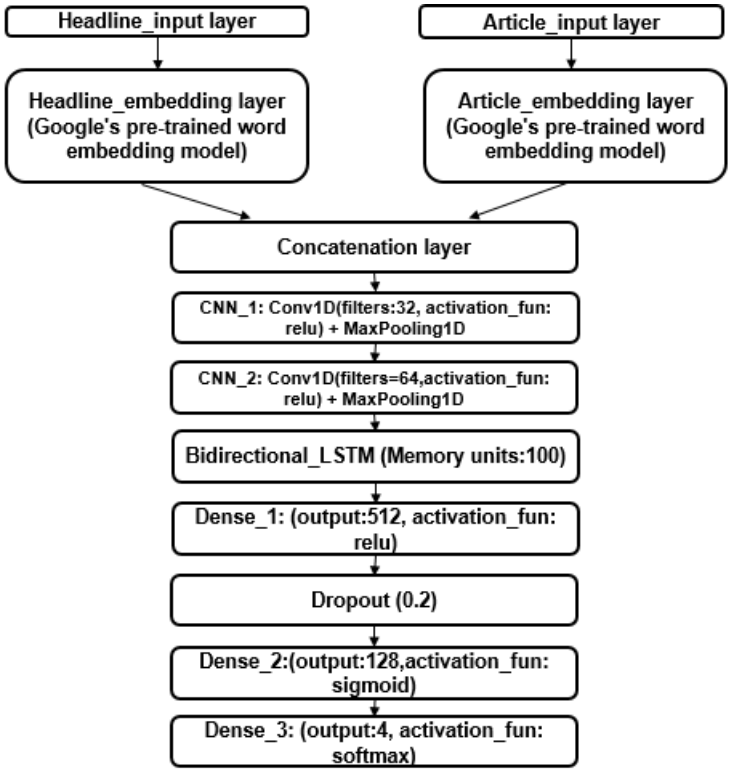


Fig. 2: Model1 Diagram; FND_Bidirectional LSTM concatenated

The output of the bidirectional LSTM passed through three dense layers (512, 128, 4 units) separated by one dropout. Using soft-max activation function in the output layer, the result is a stance classification (unrelated, agree, disagree, or discuss). Through the training process, the loss function used was sparse_categorical_cross-entropy[6] with rmsprop optimizer.

FND_Multi-head LSTM model: In this model, the merging of headline and article texts occured before passing them as input through the input layer. Then, the embedding layer using Google's pre-trained model is performed. The embedding layer resulted from previous step is passed through two CNN's layers with 32 and 64 filters, respectively that are shown in Figure 3. This step is followed by max pooling to avoid over-fitting. Then, the output is passed through five Multi-Head LSTM layers[7] with 150 memory units. A flatten layer is added, then the result are passed through one dense layer with 4 units and a softmax activation function to classify the stance into four classes (unrelated, agree, disagree, or discuss). The loss function in the training process was sparse_categorical_cross-entropy with Adam optimizer [27].

Finally, we need to mention that the number of layers and units have been chosen carefully after applying several experiments to make a decision.

[6]losses, https://keras.io/losses/

[7]keras-multi-head, https://pypi.org/project/keras-multi-head/

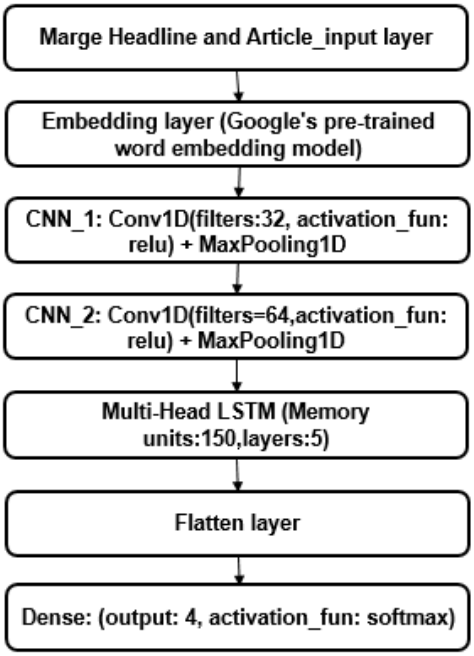


Fig. 3: Model2 Diagram; FND_Multi-head LSTM

V. Results

A. Models Evaluation

For our experiment, we have used the online python environment (Colab), which is provided by Google[8]. Tuning the hyper-parameters to select the best values and exchanging the loss and optimizer function were very challenging in this project.

Table IV illustrates a set of statistics (precision, recall, F1 scores, and accuracy) for our two models. As we have mentioned earlier, Model1 (FND_Bidirectional LSTM concatenated) has the best results. Moreover, Figure 4a and 4b show the loss function and accuracy for training and validation sets overtime for model1. The confusion matrix for model1 is also shown in Table V.

B. Discussion

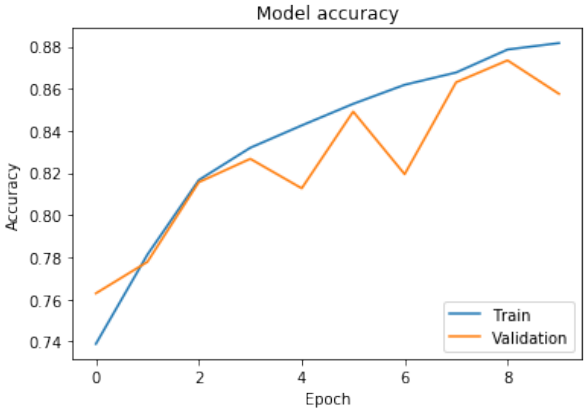
As we have mentioned earlier, we used a dateset that is provided by FNC-1 challenge. This article [13] provided a deep analysis for the first three top-performing systems in the challenge [9]. Other researchers invested their time on generating new fake news detecting techniques based on the challenge rules and score computation [21, 22]. We have used different splitting point for training, validation, testing dataset. In our experiments, we divided the dataset into train, valdiation and test sets with 60%, 20%, 20% percentage of dataset, respectively. In [22], they divided the dataset into 80%, 20% as train and test data. Furthermore, in [21], the complete training set consisted of 94% of dataset and they left the reminder to the development set for performance evaluation purposes. As a result, it is hard

[8]Google Colab, https://colab.research.google.com/

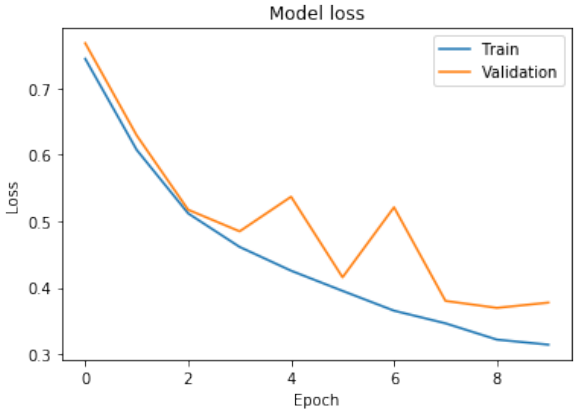
[9]http://www.fakenewschallenge.org/#fnc1results

TABLE IV: Models statistics

Models	Precision	Recall	F1_score	Accuracy
Model1: FND_Bidirectional LSTM concatenated	0.7711	0.7561	0.7635	**0.853**
Model2: FND_Multi-head LSTM	0.8451	0.5079	0.6345	0.8294



(a)



(b)

Fig. 4: Model1 results: (a) Accuracy; and, (b) Loss;

to compare the performance of the used techniques for the existing work as there is no standard test data that were used in other existing work to compare with.

TABLE V: Model1 Confusion Matrix

	Unrelated	Disagree	Agree	Discuss
Unrelated	6711	0	100	502
Disagree	67	1	107	9
Agree	255	1	406	43
Discuss	322	0	53	1408

VI. Conclusion

The evolving of new trends in computing (mobile cloud computing, wiFi, and smart devices applications) led to the widespread of internet usage. As a result, sharing information (text, audio, and video) is getting easier. Recently, it is noticed that there is a significant role of using different online platforms to share false data and spreading fake news to serve several purposes, manipulate the stock market, or just spread unwanted emotions among the online platforms users such as anger and hate. Based on that, there is an increasing demand for an accurate and efficient fake news automated detection systems. In this research, we investigated the automatic identification of fake news over online communication platforms. This task is considered a challenge for the existing content-based analysis of traditional methods. We proposed an automatic identification of fake news model using modern Machine Learning techniques, mainly deep learning and neural networks. We explored in details two models, namely, Bidirectional LSTM concatenated, and Multi-head LSTM. These models were applied to the FNC-1 dataset and the results showed that the Bidirectional LSTM concatenated model has the highest accuracy with 85% followed by the Multi-head LSTM model of about 83% accuracy. In terms of precision, the LSTM model has the highest precision of 88% followed by a Multi-head LSTM model with 84.5% precision. Overall, we recommend using the Multi-head LSTM model since it provides high precision and accuracy.

References

[1] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.

[2] J. Gottfried and E. Shearer, *News Use Across Social Medial Platforms 2016*. Pew Research Center, 2016.

[3] D. M. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild *et al.*, "The science of fake news," *Science*, vol. 359, no. 6380, pp. 1094–1096, 2018.

[4] A. Bovet and H. A. Makse, "Influence of fake news in twitter during the 2016 us presidential election," *Nature communications*, vol. 10, no. 1, p. 7, 2019.

[5] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," *Proceedings of the Association for Information Science and Technology*, vol. 52, no. 1, pp. 1–4, 2015.

[6] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, "Novel visual and statistical image features for microblogs news verification," *IEEE transactions on multimedia*, vol. 19, no. 3, pp. 598–608, 2017.

[7] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha, "Detecting rumors from microblogs with recurrent neural networks." in *Ijcai*, 2016, pp. 3818–3824.

[8] N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news detection," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 2017, pp. 797–806.

[9] X. Zhou and R. Zafarani, "Fake news: A survey of research, detection methods, and opportunities," *arXiv preprint arXiv:1812.00315*, 2018.

[10] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22–36, 2017.

[11] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," in *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 3391–3401.

[12] R. Davis and C. Proctor, "Fake news, real consequences: Recruiting neural networks for the fight against fake news," 2017.

[13] K. Miller and A. Oswalt, "Fake news headline classification using neural networks with attention," tech. rep., California State University, year, Tech. Rep., 2017.

[14] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, and J. Gao, "Eann: Event adversarial neural networks for multi-modal fake news detection," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2018, pp. 849–857.

[15] J. Zhang, L. Cui, Y. Fu, and F. B. Gouza, "Fake news detection with deep diffusive network model," *arXiv preprint arXiv:1805.08751*, 2018.

[16] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake news detection on social media using geometric deep learning," *arXiv preprint arXiv:1902.06673*, 2019.

[17] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *Journal of economic perspectives*, vol. 31, no. 2, pp. 211–36, 2017.

[18] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," *Proceedings of the Association for Information Science and Technology*, vol. 52, no. 1, pp. 1–4, 2015.

[19] D. Klein and J. Wueller, "Fake news: A legal perspective," *Journal of Internet Law (Apr. 2017)*, 2017.

[20] A. Hanselowski, P. Avinesh, B. Schiller, F. Caspelherr, D. Chaudhuri, C. M. Meyer, and I. Gurevych, "A retrospective analysis of the fake news challenge stance-detection task," in *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 1859–1874.

[21] A. K. Chaudhry, D. Baker, and P. Thun-Hohenstein, "Stance detection for the fake news challenge: identifying textual relationships with deep neural nets," *CS224n: Natural Language Processing with Deep Learning*, 2017.

[22] S. Chopra, S. Jain, and J. M. Sholar, "Towards automatic identification of fake news: Headline-article stance detection with lstm attention models," 2017.

[23] N. Lozhnikov, L. Derczynski, and M. Mazzara, "Stance prediction for russian: data and analysis," in *International Conference in Software Engineering for Defence Applications*. Springer, 2018, pp. 176–186.

[24] I. Augenstein, T. Rocktäschel, A. Vlachos, and K. Bontcheva, "Stance detection with bidirectional conditional encoding," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016, pp. 876–885.

[25] W. Ferreira and A. Vlachos, "Emergent: a novel data-set for stance classification," in *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies*, 2016, pp. 1163–1168.

[26] L. Ma and Y. Zhang, "Using word2vec to process big text data," in *2015 IEEE International Conference on Big Data (Big Data)*. IEEE, 2015, pp. 2895–2897.

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *The International Conference on Learning Representations (ICLR)*, 2015.