Taylor & Francis
Taylor & Francis Group

Applied Artificial Intelligence
An International Journal

ISSN: 0883-9514 (Print) 1087-6545 (Online) Journal homepage: https://www.tandfonline.com/loi/uaai20

# Fake News Identification and Classification Using DSSM and Improved Recurrent Neural Network Classifier

Shrutika S. Jadhav & Sudeep D. Thepade

**To cite this article:** Shrutika S. Jadhav & Sudeep D. Thepade (2019) Fake News Identification and Classification Using DSSM and Improved Recurrent Neural Network Classifier, Applied Artificial Intelligence, 33:12, 1058-1068, DOI: 10.1080/08839514.2019.1661579

**To link to this article:** https://doi.org/10.1080/08839514.2019.1661579

APPLIED ARTIFICIAL INTELLIGENCE
2019, VOL. 33, NO. 12, 1058–1068
https://doi.org/10.1080/08839514.2019.1661579

Taylor & Francis
Taylor & Francis Group



# Fake News Identification and Classification Using DSSM and Improved Recurrent Neural Network Classifier

Shrutika S. Jadhav and Sudeep D. Thepade

Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India

**ABSTRACT**

The widespread use of social media has enormous consequences for the society, culture and business with potentially positive and negative effects. As online social networks are increasingly used for dissemination of information, at the same they are also becoming a medium for the spread of fake news for various commercial and political purposes. Technologies such as Artificial Intelligence (AI) and Natural Language Processing (NLP) tools offer great promise for researchers to build systems, which could automatically detect fake news. However, detecting fake news is a challenging task to accomplish as it requires models to summarize the news and compare it to the actual news in order to classify it as fake. This project proposes a framework that detects and classifies fake news messages using improved Recurrent Neural Networks and Deep Structured Semantic Model. The proposed approach intuitively identifies important features associated with fake news without previous domain knowledge while achieving accuracy 99%. The performance analysis method used for the proposed system is based on accuracy, specificity and sensitivity.

## Introduction

Nowadays Internet and social media have made the access to the news information much easier and comfortable. Often Internet users can follow the events of their interest in online mode, and access via mobile devices makes this process even easier. But with great possibilities come great challenges. Social media have a huge influence on the society, and as it often happens, there is someone who wants to take advantage of this fact. Sometimes to achieve some goals social-media may manipulate the information in different ways. This leads to producing of the news that are not completely true or even completely false (Campan, Cuzzocrea, and Truta 2017; Tschiatschek et al. 2018). Twitter is one of the widely used social media platform by all generations and fields. Because of its widespread reach of

**CONTACT** Shrutika S. Jadhav shrutikajadhav0595@gmail.com Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uaai.

more than 330 million users, it contains all forms of data on its website. As a result, it is also considered to be the newspaper of the 21st century because of the ability of every user to report major events and incidents worldwide. Through tweets, people share articles, videos and photos where almost 85% of the topics discussed on Twitter relate to the news (Smart Insights 2019). Twitter is a free micro-blogging platform for social networking that allows users to broadcast tweets. The re-tweeting mechanism is also effective in redistribution of the tweets as well as hashtags and trend topics help people search for the events. This freedom for the user to post anything also results in the spread of fake information. People are also less likely to check the news shared by their friends and thus spread fake news more quickly.

Fake news is a type of yellow journalism or propaganda that consists of misinformation. It has traditionally been spread through print and broadcast mediums, but with the rise of social media, it can now be disseminated virally. Fake news detection is the task of classifying news according to its veracity. In a simple setting, it is a binary classification task, whereas in a more difficult setting it is a fine grained classification task. Fake news detection is one of the hottest research topics in artificial intelligence recently. Due to the availability of the Internet and the readiness of sharing information through social media, it is easy to make fake news and propagate it worldwide. When being spread widely, the fake news may cause substantial adverse impact to many aspects of life. Consequently, an amount of research has been conducting recently to spot fake news. Despite receiving significant attention from the crowd, fake news detection did not gain much improvement for a while due to the insufficiency of fake news data. In fact, it was not until 2017 with the introduction of Liar dataset, has fake news detection shown some noticeable performance (Shu et al. 2017; Wang 2017).

Many scientists believe that fake news issue may be addressed by means of data mining algorithms and deep learning models. As recently data mining algorithms have started to work much better on lots of classification problems (image recognition, voice detection and so on) as hardware's are cheaper and bigger datasets are available easily. The major drawback of these baseline classifiers is that it does not capture semantics in text and semantic features would be very useful for fake news identification. Fake news has potential negative impact on society in order to minimize the harmful effect of fake news, there is a need to build an improved model. Deep Learning is widely used for the development of decision support model based on expert knowledge and observations. Fake news problem can be tackled using deep learning models. Deep Neural Network (DNN) is a network with multiple hidden layers between the input and output layers. It can model complex nonlinear relationships as well as capture semantics in text. Deep Neural Networks are trained using new methods to outperform traditional methods of identification of fake news. Our system describes

a simple fake news detection method based on deep learning algorithms like improved RNN and DSSM. The goal of the research is to examine how this particular method works for this problem with provided news dataset and to support the idea of using deep learning model for fake news detection.

## Related Work

The proliferation of fake news on social media has become a major problem in our society. Identification of such fake news is a necessary task to avoid its negative effects. In Parikh and Atrey (2018), authors described existing fake news detection approaches that are based on text-based analysis along with some popular fake news datasets. In Castillo, Mendoza, and Poblete (2011), they developed an automatic method for assessing the credibility of a given set of tweets. The methods used were evaluated by a significant number of human assessments about the credibility of items on recent sample of twitter postings.

In Granik and Mesyura (2017), the authors proposed a simple approach for fake news detection using Naive Bayes classifier, which showed that simple artificial algorithm may give a good result on such important problem. Considering the work of baseline classifiers in Gupta and Kaushal (2015), authors has proposed a novel method that combines advantages of three learning algorithms namely Naive Bayes, Clustering and Decision trees. Individual algorithms were compared with combined approach and was observed that combine approach could give best results.

In Wu and Liu (2018), the authors proposed a CSI model which combines three modules for accurate and automatic prediction. Specifically, they incorporated the behavior of both parties, users and articles, and the group behavior of users who propagate fake news. Motivated by the three characteristics a model called CSI is built which is composed of three modules: Capture, Score, and Integrate. The first module uses a Recurrent Neural Network to capture the temporal pattern of user activity on a given article which is based on the response and text. The second module learns the source characteristic considering behavior of users, and both modules are integrated with the third module to classify an article as fake or not. In Ruchansky, Seo, and Liu (2017), Liu and Wu (2018), a method is proposed that detects and classifies fake news messages from Twitter posts using hybrid of convolutional neural networks and long-short term recurrent neural network models. The hybrid implementation of two deep learning models was done to find out that the large numbers of tweets about an event to determine the veracity or credibility of the messages are not required. This approach gave a boost in the achievement of a higher performance while not requiring a large amount of training data typically associated with deep

**Table 1.** Performance comparison of classifier.

| Methodology | Dataset | Accuracy |
|---|---|---|
| Naive Bayes Algorithm (Granik and Mesyura 2017) | BuzzFeed News Dataset | 74% |
| Recurrent Neural Network (Wu and Liu 2018) | Twitter Dataset | 89.2% |
| Convolutional Neural Network – LSTM (Ruchansky, Seo, and Liu 2017) | Pheme Dataset (Textual and Visual) | 82% |
| DSSM-LSTM | LIAR Dataset | 99% |

learning models. The proposed work using this deep learning approach has achieved 82% accuracy.

## Current State of Art in Fake News Detection

Identifying fake news is a complex task due to characteristics of fake news. The news identification strategies utilize multiple news-related and social-related types of information. Recently fake news detection systems mainly targets four prospective of knowledge, style, propagation and credibility based on which news are identified (Zhou and Zaarani 2018). Nevertheless, in early 2017 a considerable number of worldwide entities, organizations and researchers aimed at stopping of fake news that included techniques like human intervention and using algorithms to verify information veracity. Considering these techniques automatic methods are developed that assessed credibility on twitter with significant number of human assessment. Systems using classification algorithms are proposed that exploit various simple artificial algorithms like Naive Bayes, clustering and decision trees. Deep Learning approach build more effective models for fake news problem that considered visual and textual dataset as well as semantic features in text which were not captured by baseline classifiers. The following Table 1 shows the comparison based on accuracy for different methodology used on various datasets.

## Proposed System Architecture

By considering drawbacks of existing system, we proposed a reliable system. This proposed system helps to identify fake news existing in LIAR dataset.

Figure 1. showing the generic working of proposed system in block diagram. In this work, we are discussing the effectiveness of combination of DSSM and improved RNN architectures for the detection of fake news using the twitter dataset. In this hybrid model, DSSM consists of word hashing and multi-layer nonlinear projection. The output of the multi-layer nonlinear projection layer is fed to the subsequent improved RNN layer to learn the long-range temporal dependencies.

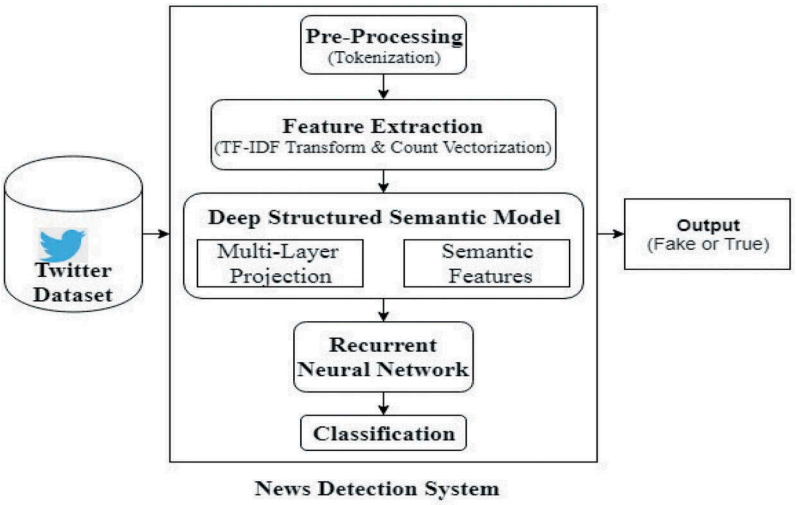


Figure 1. Block diagram of proposed system.

## System Analysis

### Methodology

The proposed framework detects and classifies fake news using DSSM and improved RNN model from twitter dataset. In this model, the input strings from the twitter dataset are fed into the input layer. These strings are pre-processed to form bag of words. These bags of words are further passed to DSSM model. In DSSM model three steps are performed where first step is word hashing, second step is multi-layer nonlinear projection and third step is semantic feature generation. These initial layers of the DSSM architecture represent the feature extraction mechanism while the improved RNN layers represent the classification mechanism. The detailed framework of all the layers, which define the structure of the neural network is explained below:

### Input Layer

The LIAR dataset is the input provided to the input layer. This dataset consists of titles and text of various news articles.

### Preprocessing and Feature Extraction

Preprocessing transforms the data into a format that will be more easily and effectively processed for the purpose of the user. The input twitter dataset contains data in textual format which is preprocessed using tokenization and further features are extracted using TF-IDF transform (Buntain and Golbeck 2017) and count vectorization.

### Deep Semantic Structural Model

The input (raw text features) to the DSSM is a high dimensional term vector, for example, raw counts of terms in a query or a document without normalization, and the output of the DSSM is a concept vector in a low-dimensional semantic feature space.

The typical DSSM architecture developed here for mapping the raw text features into the features in a semantic space is shown above in Figure 2. The word hashing method described here aim to reduce the dimensionality of the bag-of-words term vectors. It is based on letter n-gram, and is a new method developed especially for our task. Given a word (e.g. good), we first add word starting and ending marks to the word (e.g. #good#). Then, we break the word into letter n-grams (e.g. letter trigrams: #go, goo, ood, od#). Finally, the word is represented using a vector of letter n-grams. Multilayer nonlinear projection identifies keywords/concepts further generating semantic features.

$$R(S,T) = \mathrm{Cos}\left(y_{s,}y_t\right) = \frac{y_s^T, y_t}{||y_s||||y_t||} \tag{1}$$

Similarity between source and target vector is computed using above cosine similarity Equation (1).

### Improved RNN

LSTM is an improved form of Recurrent Neural Network (RNN) (Dimpas, Po, and Sabellano 2017). LSTM introduces memory blocks instead of conventional simple RNN units to handle the problem of vanishing and exploding gradient. LSTMs can handle long term dependencies much better than the traditional RNNs. This means that LSTMs can remember and connect previous information to the present.

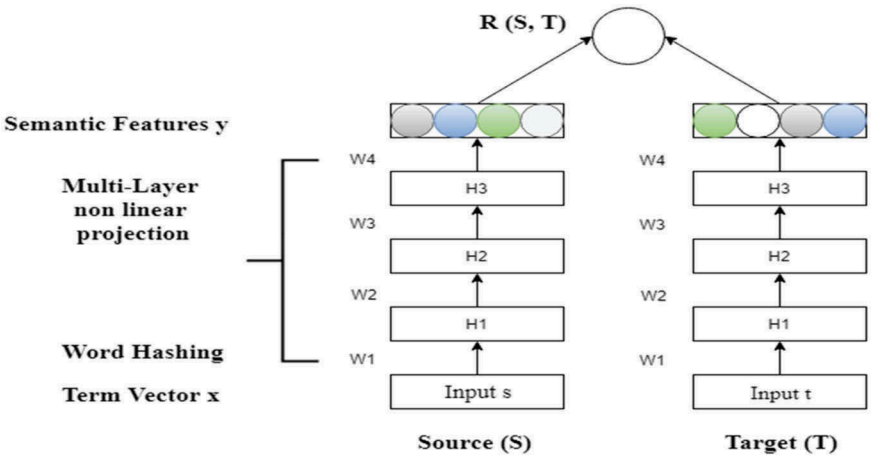


Figure 2. DSSM architecture.

Thus final layer of the model is the LSTM layer which gives the classified output. The DSSM layers are used to extract features form the input data while the LSTM layer is used for prediction and classification. This layer consists of four gates which are defined as:

Input activation:

$$a_t = \tanh(W_a \cdot x_t + U_a \cdot out_{t-1} + b_a) \tag{2}$$

Input gate:

$$i_t = \sigma(W_i \cdot x_t + U_i \cdot out_{t-1} + b_i) \tag{3}$$

Forget gate:

$$f_t = \sigma(W_f \cdot x_t + U_f \cdot out_{t-1} + b_f) \tag{4}$$

Output gate:

$$o_t = \sigma(W_o \cdot x_t + U_o \cdot out_{t-1} + b_o) \tag{5}$$

which lead to:
Internal state:

$$state_t = a_t \Theta\ i_t + f_t \Theta state_{t-1}) \tag{6}$$

Output:

$$out_t = \tanh(state_t) \Theta\ o_t \tag{7}$$

It maps the features extracted from the previous layer to their appropriate class. An activation function called the softmax function is used to classify these extracted features and to predict the output.

### *Output Layer*

The output obtained is in binary form. The output layer provides output values 0 (fake news predicted yes) or 1 (fake news predicted no).

## Flow of Proposed System

This proposed system helps to identify fake news existing in twitter dataset using the described flow of proposed system from Figure 3.

The proposed system is divided into two phases:
Training and Testing

(1) Steps of Training phase
- Insert Training dataset.
- Preprocess this data using tokenization and stemming methods.
- Extract Features using TF-IDF transform.
- Generate DSSM-LSTM model and obtain an object file.

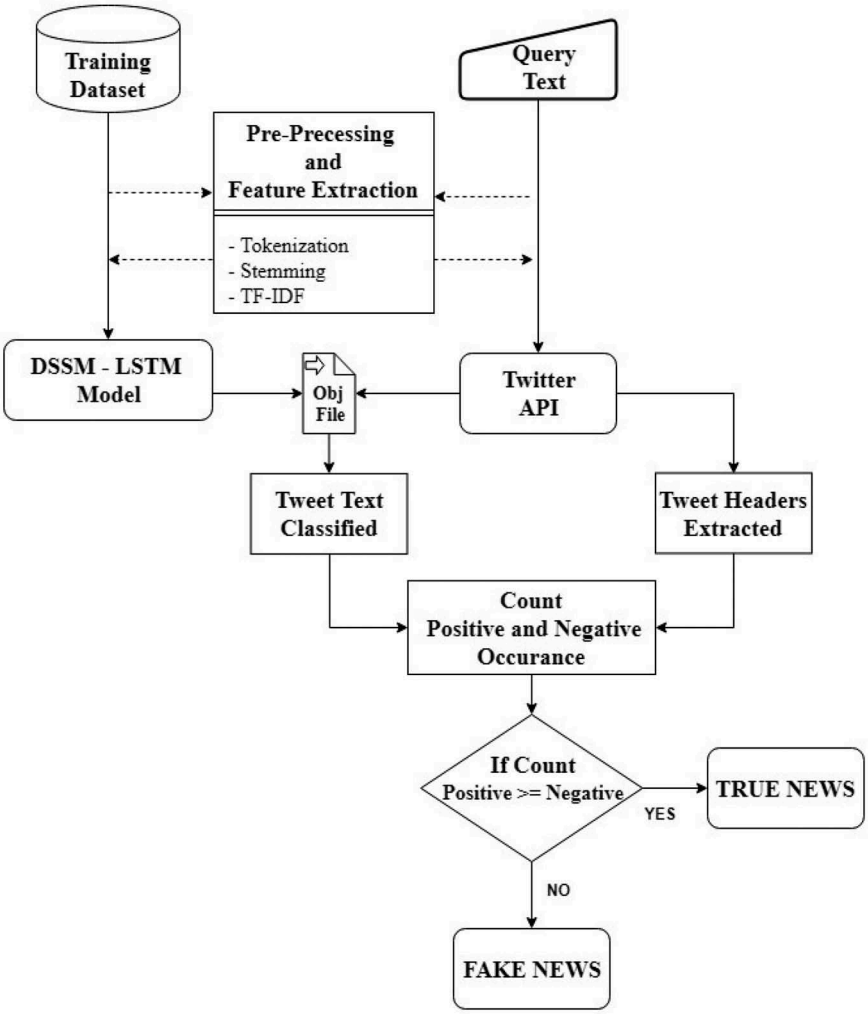


**Figure 3.** Flow of proposed system.

(2) Steps of Testing phase

- Insert Text Query from Test Dataset.
- Preprocess each text query using tokenization and stemming method.
- Extract Features using TF-IDF transform and pass them to twitter API.
- Generated twitter data is passed to obj file.
- Classified tweet text from obj file is used to count the positive and negative occurrences of tweet header.
- News is verified based on the count of positive and negative count. (positive ≥ negative then news is positive and vice versa)

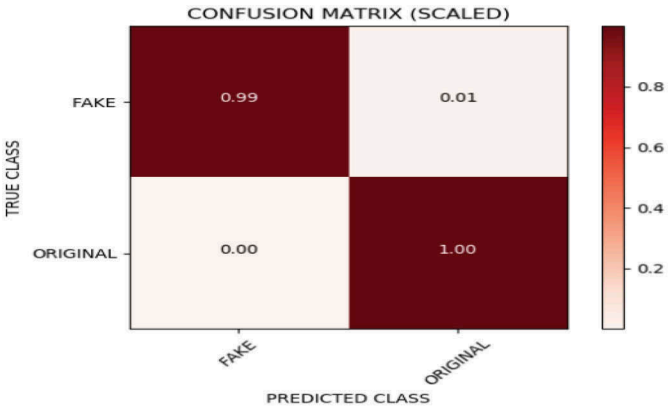


**Figure 4.** Confusion matrix.

## Results and Discussions

The proposed system is implemented in python using PyCharm IDE and Windows XP/7 operating system. The hardware requirements are as mentioned—Processor: Pentium 4.0 or higher, RAM: 3 GB or higher RAM, Storage: 40 GB or higher, System type: 64-bit/32-bit operating system.

The input twitter dataset are preprocessed and converted into bag of words. The bags of words are then passed to DSSM-LSTM model for classification. The results are obtained in the form of confusion matrix which evaluates the performance of the system for fake news detection problem. The confusion matrix is shown in Figure 4 and from the matrix performance measures such as accuracy, specificity and sensitivity are calculated.

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{8}$$

$$\text{Specificity} = \frac{TN}{FP + TN} \tag{9}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{10}$$

Considering the equation of performance measures and confusion matrix we obtain accuracy as 99%, specificity as 99% and sensitivity as 100%. The system results are compared with the results obtained by implementing both algorithms separately. The following table shows the comparison of classifiers for three different data splits, that is, 50–50%, 80–20% and 75–25% based on accuracy.

The following graph in Figure 5 shows the overall comparison of hybrid DSSM-LSTM model with individual DSSM and LSTM based on accuracy.

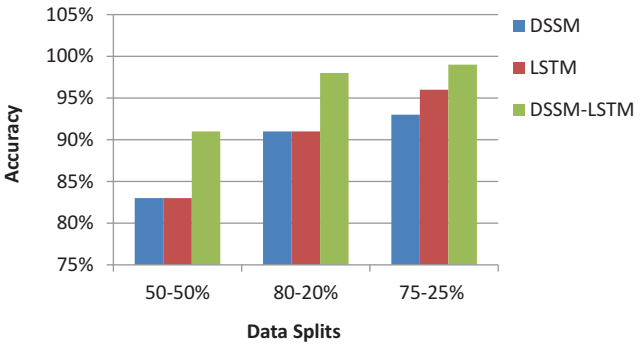


**Figure 5.** Performance comparison of classifiers.

**Table 2.** Performance comparison of classifier based on accuracy.

| Classifier | 50–50% | 80–20% | 75–25% |
|---|---|---|---|
| DSSM | 83% | 91% | 93% |
| LSTM | 83% | 91% | 96% |
| DSSM-LSTM | 91% | 98% | 99% |

We conclude from the Table 2 and Figure 5 the values of DSSM-LSTM model with 75–25% data split have greater performance than other classifiers having different splits.

## Conclusion

Many people follow social media news instead of traditional media. Social media has also been used to spread false news that has negative consequences on individual people and society. In this paper, an innovative DSSM and improved RNN model for fake news detection has been presented with detail description. The results obtained describes that the DSSM-LSTM model have greater performance than the other classifiers depending on accuracy performance measure parameter. This model takes news events as an input and based on twitter reviews and classification algorithms it predicts news being fake or real with the accuracy as 99%. Hence the proposed work is highly desirable to classify fake news and to increase the accuracy.

We have focused on the twitter social media. Thus, our future work would be to integrate other social media in our system.

## References

Buntain, C., and J. Golbeck. 2017. Automatically identifying fake news in popular twitter threads. 2017 IEEE International Conference on Smart Cloud, SmartCloud 2017, New York City, NY, USA, November 3-5, 2017.

Campan, A., A. Cuzzocrea, and T. M. Truta. 2017. Fighting fake news spread in online social networks: Actual trends and future research directions. 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, December 11-14, 2017.

Castillo, C., M. Mendoza, and B. Poblete. 2011. Information credibility on twitter. Proceedings of the 20th international conference on World wide web, ACM, Hyderabad, India, March 28 - April 1, 2011.

Dimpas, P. K., R. V. Po, and M. J. Sabellano. 2017. Filipino and English clickbait detection using a long short term memory recurrent neural network. 2017 International Conference on Asian Language Processing, IALP 2017, Singapore, December 5-7, 2017.

Granik, M., and V. Mesyura. 2017. Fake news detection using naive Bayes classifier. 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kyiv (*Ukraine*), May 29 – June 2, 2017.

Gupta, A., and R. Kaushal. 2015. Improving spam detection in online social networks. 2015 International Conference on Cognitive Computing and Information Processing (CCIP), Noida, India, March 3-4, 2015.

Smart Insights. 2019. Global social media research summary 2019| Smart insights. https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/.

Liu, Y., and Y.-F. B. Wu. 2018. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. Thirty-Second AAAI Conference on Artificial Intelligence, Hilton New Orleans Riverside, New Orleans, Louisiana, USA, February 2–7, 2018.

Parikh, S. B., and P. K. Atrey. 2018. Media-rich fake news detection: A survey. 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Miami, Florida, USA, April 10-12, 2018.

Ruchansky, N., S. Seo, and Y. Liu. 2017. Csi: A hybrid deep model for fake news detection. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017.

Shu, K., A. Sliva, S. Wang, J. Tang, and H. Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter* 19 (1). doi:10.1145/3137597.3137600.

Tschiatschek, S., A. Singla, M. G. Rodriguez, A. Merchant, and A. Krause. 2018. Fake news detection in social networks via crowd signals. International World Wide Web Conferences Steering Committee, Lyon, France, April 23 - 27, 2018.

Wang, W. Y. 2017. "liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv Preprint arXiv:1705.00648*.

Wu, L., and H. Liu. 2018. Tracing fake-news footprints: Characterizing social media messages by how they propagate. Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM 2018, Marina Del Rey, CA, USA, February 5-9, 2018.

Zhou, X., and R. Zaarani. 2018. Fake news: A survey of research, detection methods and opportunities. *arXiv Preprint arXiv:1812.00315*.