# Misreporting and Fake News Detection Techniques on the Social Media Platform

Yuran Feng

Department Of Mathematics, College of Arts and Sciences, University of Washington-Seattle, WA, United States

yuranf2@uw.edu

**Abstract.** One of the major concerns nowadays is the rapid spreading of fake news or unverified information on all kinds of social media. Misinformation and disinformation on the digital media of news distribution have brought significant negative impacts to our community, which the traditional techniques can no longer detect and deal with it effectively. It is urgent to squelch fake news immediately to limit its impact on the economy and society. As deep learning techniques continue developing in recent decades, scholars successfully deployed deep neural networks on fake news detection tasks. The first noticeable thing is to admit that the fake news detection task has made significant accomplishments as fast as we hoped. It is necessary to study further and broadly review the state-of-the-art fake news detection approaches. In this review paper, we survey several distinct deep learning techniques and provide a comprehensive review of automatic fake news detection classification tasks and the datasets and models used, demonstrating the performance evaluation on different approaches. We have also analyzed the potential challenge we encountered in fake news detection.

**Keywords:** Fake news, Machine learning, Deep learning, Text classification, Social media, Natural language processing.

## 1. Introduction

The proliferation of the Internet has promoted the development of online social media as it provides a fast and convenient way to disseminate information, and the traditional media no longer satisfy people's daily consumption of news nowadays. As the Internet becomes more embedded in every aspect of our daily lives, people tend to spend more time on digital media instead of traditional media (i.e., newspapers). The digital media of news distribution has become our main course of action to fulfill the news consumption. Since the primary media channel has shifted to online, the interaction between self (individual) and groups or interaction between each group are more reliant on social networks, such as Facebook, Instagram, and Twitter. For example, according to [1], the number of accounts on Twitter, one of the most popular micro-blogging platforms, has risen from 1.6 million to 32.1 million. The popularity of social media even led the U.S. State Department to intervene in the routine maintenance of Twitter during the Iranian poll since Iranians used it to coordinate protests [2].

The authenticity of the news content is a significant factor of concern nowadays because most people on earth rely on social media to perceive trending topics. When social media become a powerful means of information, it also negatively impacts society by spreading disinformation and fake news, which influences the public's views. A catchy headline can increase the clickthrough rate, leading to an article being shared thousands of times. For example, the survey in [3] suggests that, during the presidential election in the U.S. in 2016, people shared fake news that had the headline "Pope Francis had endorsed Donald Trump's presidential candidacy" more than one million times on Facebook, and some people in the survey reported believing the headline. The absence of fact-checking over social media posts makes it a perfect fertile land for spreading fake news and unverified information. Everyone can quickly post and disseminate disinformation over the Internet. As a result, social media platforms and researchers devote themselves to identifying and squelching the questionable content to be false statements. As stated by [4], Facebook is already processing identifying fake content by flagging unverified news, with the help of users, as "disputed by third-party fact-checkers," which helps users avoid spreading fake news by informing them that the post is

"disputed by third parties" before they accidentally share it. The identification of Fake news on the social platform is a challenging task that often requires human involvement. This traditional approach that identifies fake news by human overlook the volume of news.

Researchers in the academic world, on the other hand, strive to deploy modern techniques on this task for efficiency and timely. Machine learning (ML) is initially used as an approach for fake news detection as it evaluates the credibility of sources and assesses the truthfulness of claims in the news based on feature extraction [5]. For instance, SVMs have supervised machine learning algorithms that employ discriminative classifiers, defined by separating hyperplanes in N-dimensional space. Therefore, the data points that fall on either side of the hyperplane are assigned to different classes. Although SVMs performed well for deception detection in text with an F-measure of 0.84 from the experiment in [6], the performance varies greatly in the different datasets we selected for training.

These machine learning techniques heavily depend on manually crafted features from the content, user profile, and different characteristics of posts, and feature extraction is time-consuming, often biased, and labor-intensive [7]. The feature extraction process by human intervention requires an expert with strong background knowledge in the relevant area. These features could only represent the temporal characteristic of fake news and disinformation. [7] demonstrates two-time series plots that depict the classic shallow patterns of rumor and non-rumor signals from [8]. The difference between the plots in [7] is not clear-cut and strong, which is not suitable for feature engineering.

On the other hand, in different detection approaches, Deep Learning (DL) has been proven to have great advantages [9] in automated fake news detection in research communities. The deep neural networks are able to learn hidden representations and capture the relevant context details through multi-nonlinear transformation. The recurrent neural networks (RNNs) are suitable for modeling sequential data in which it exploits deep data representation because the connections between each node form a direct cycle, which eventually creates an internal state of the network [10]. This internal structure allows RNNs to capture both textual and temporal characteristics from news content [7]. Because of the massive potential of deep learning in natural language processing [10], more and more scholars in research communities employ deep learning in Fake news detection.

There have been many research papers about applying deep neural networks to fake news identification. Academics have also achieved high scores on testing the accuracy performance and made significant improvements in the generalizability of algorithms and semantic content. Time up to today, the research communities still demand more comprehensive and effective surveys on fake news recognition. Many people who are new to the field of deep learning lack knowledge regarding the state-of-the-art approaches in deep learning and how these techniques can be applied to the real industry to accomplish fake news detection. This survey includes recent academic works that further assist beginners in briefly acknowledging the fake news detection method. Hence, the main contribution of this survey is to provide a comprehensive review of automatic fake news detection and a summary of the current industry developments in the field of deep learning, encouraging subsequent scholars to make suggestions and improve the performance of contemporary technologies in the industry.

The rest of the paper is organized as follows. Section 2 describes fake news detection and distinguishes the relevant deep learning application scenarios under different situations. It further provides relevant datasets and evaluation metrics. Section 3 elaborates on the mainstream deep learning from the experimental studies in several papers and describes these state-of-the-art methods in detail. Section 4 reflects the imperfection and challenges of fake news detection with the deep neural network. Finally, Section 5 concludes the paper.

## 2. Background

In this section, we discuss how to evaluate the performance of the ML or DL model. We provide specific details of this task and its available datasets and evaluation metrics.

### 2.1. Task

Fake news is defined as a narrative made up of no verifiable evidence or references and may be propaganda intended to deceive the reader. Therefore, here we list some fake news detection methods based on classification approaches. The machine learning algorithms are very effective for solving tasks in information engineering. Most machine learning models implement supervised learning strategies to make predictions, such as regression, random forest, SVM, and decision tree. Deep learning is also one of the most trending research focus areas in fake news detection. The two most widely used neural networks are Recurrent Neural networks (RNN) and Convolutional Neural networks (CNN).

### 2.2. Datasets

A massive labeled benchmark dataset is essential for DL and ML models to properly learn the weights during the training process. Datasets can differ in size, label, and domain of data. The following datasets are used in different studies for training and testing purposes. There are mainly two types of datasets we covered in fake news detection.

LIAR is a publicly large-scale dataset for fake news detection. This dataset includes 12.8K short statement claims from POLITIFACT.COM's API. Each statement is manually labeled and annotated as true/false by a POLITIFACT.COM editor. FNC-1 focuses on the task of stance detection that involves estimating the relative perspective of two-piece of content, and it also extends the work of Ferreira & Vlachos [20]. This dataset contains 75385 labeled pairs in which each pair contains a headline and an article. The body of each article is labeled as either unrelated, discuss, disagree, or agree relative to a headline. RealNews is also a news article dataset scraped from Common Crawl with newspaper python library. The training sample includes the news dump from Common Crawl ranging from December 2016 to March 2019, and the articles in April 2019 dump are used for evaluation purposes. Table 1 shows the details of the datasets.

The Weibo dataset is a Chinese Named Entity Recognition dataset taken from Sina Weibo, a Chinese social media platform. The dataset includes user preference, region, organization, and user posting. UPFA dataset can evaluate binary graph classification, graph anomaly detection, and fake news detection tasks. The special thing about this dataset is that it includes fake news propagation networks(retweet), originally extracted by FakeNewsNet, on Twitter. Each news retweet graph is represented by a tree where the root represents news, and the corresponding leaf nodes represent all users who retweeted the news. Table 1 shows the details of the datasets.

**Table 1.** Publicly available datasets

| Datasets | Size | Modality | Labels | Type |
|---|---|---|---|---|
| LIAR [11] | 12.8K | Text | Pants on fire, false, barely true, half-true, mostly true, and true | Political statements |
| Weibo [12] | 40k tweets | Text & image | Unverified, true, false, non-rumor | Social media data |
| FNC-1 [13] | 75K | Text | Agrees, disagrees, discusses, unrelated | News articles |
| FakeNewsNet [14] | 5K | Text | Fake, real | News articles, social media data |
| MuMiN [15] | 21M | Image & Text & graph | Fake | social media data |
| Twitter [16] | 992 | Text & image | Rumor, non-rumor | Fact-checked claims |
| UPFD [17] | 20M | Text & image | Fake | Social media data |
| RealNews [18] | 120 gigabytes | Text | - | News article |
| MM-COVID [19] | 10K | Text | Fake, Trustworthy | News article |

### 2.3. Evaluation Metrics

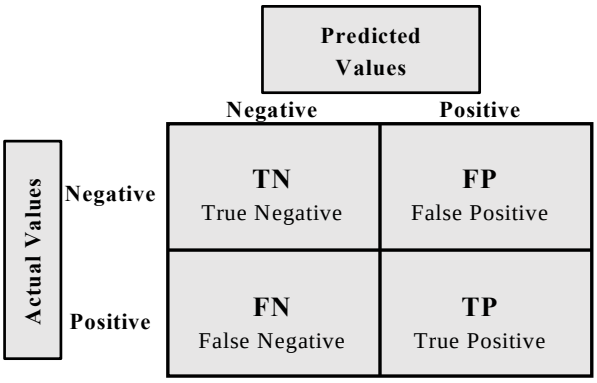



**Figure 1.** Confusion Matrix

We often use evaluation metrics to measure model performance. Before we deploy the model for production on unseen data, utilizing different evaluation metrics during the training process improves the overall predictive power of our model. To begin with, the fundamental concept for most evaluation metrics is the confusion matrix, a matrix representation of the prediction, which represents the performance of a classifier on the testing set. The prediction composes four outcomes: True Positive, True Negative, False Positive, and False Negative. The FP and FN are also called Type I and Type II errors, respectively.

A. Accuracy

The most commonly used metric in ML and DL is accuracy which can be written as the following equation:

$$\boldsymbol{Accuracy} = \frac{\#TP+\#TN}{\#FP+\#FN+\#TP+\#TN} \tag{1}$$

The classification problem primarily uses accuracy as its evaluation metric. However, the evaluation result can be unreliable when we have a class imbalance. For example, suppose we have 9999 samples for negative class and 1 sample for positive class. Then the majority class classifier achieves 99.99% accuracy by predicting all samples belonging to negative classes. Hence, it's better to consider different metrics instead of only accuracy.

B. Precision

Precision is one of the indicators of a model's performance. It represents the quality of the prediction. We define precision as the number of true positives divided by the total number of positives. Although the classifier model is able to achieve perfect precision, 100 percent, the real-world scenario also considers other factors, trading off between precision and recall.

$$\boldsymbol{Precision} = \frac{\#TP}{\#TP+\#FP} \tag{2}$$

C. Recall

The recall, also called the true positive rate of a model, indicates the frequency of predicting yes when the label is actually yes. Precision and recall together evaluate model performance in the face of class imbalance.

$$\boldsymbol{Recall} = \frac{\#TP}{\#FN+\#TP} \tag{3}$$

D. F1-Score

The harmonic mean of recall and accuracy represents the F1-score, where 1 represents the best and 0 represents the worst. It is mainly used to compare the model performance of two classifiers such that a higher F1-score score decides which one gives a better result when one classifier achieves high precision, and another achieves high recall.

$$\boldsymbol{F_1} = \mathbf{2} \times \frac{\#Precision \times \#Recall}{\#Precision+\#Recall} \tag{4}$$

## 3. Literature Survey of Recent Research

In recent decades, since online social media provides a convenient way to spread fake news and misinformation on the Internet, machine learning and deep learning researchers have focused on fake news detection. This section reviewed some recent research on detecting fake news done by different researchers using various novel or original approaches.

### 3.1. Deep Learning Approach

In [21], the author introduces an automatic fake news detection approach that utilizes a variety of features related to Facebook user accounts with news context features to examine users' behavior and detect rumors spread by users using ML and DL classifiers. The framework "chrome extension-based fake news detection approach, "which is prompted by the people passively, detects fabricated information on Facebook on the basis of analyzing fake activity collected from disseminating news and various user profiles by the crawler. The crawler imports Facebook API to pull out information on demand and stores the pre-processed information in an XLS file, which is used to construct the raw dataset and baseline dataset. The authors have studied and compared five machine learning classifiers to one deep learning classifier in a fake news classification problem, namely, Logistic Regression, K Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, Naïve Bayes, and Long Short Term Memory (LSTM). Using the cross-validation techniques for both ML and DL classifiers with ten folds, the deep learning-based analysis achieves a greater accuracy score than the ML classifier for performance evaluation on different categories of baseline dataset, videlicet, biography, news content features, and a mixture of both. The experimental observation shows that the feature combination of user profiles and news content gives the best accuracy (99.4%) for LSTM compared to other features. The study successfully established on the client-side of the chrome environment to classify fake news in the present time by crawling account profiles and sharing articles.

A recent interest of the main entity that gets the human's attention respecting trust is the article's content. Therefore, to research the influential features in linguistic contexts, the authors discussed a linguistic model in [31], to generate language-driven features. Two datasets used in [31] are Buzzfeed Political News and Random Political News, both in the Horne2017_FakeNewsData repository. It first studies the process of linguistic feature-evidence extraction based on literature for linguistic models. These feature evidence include Syntax-based evidence, Sentiment-based evidence, Grammatical evidence, and Readability-based evidence. Further, it proposed the Linguistic feature-based driven deep learning models to achieve efficient fake news detection. According to the literature study, since readability is treated as the weakest characteristic of all features, the same deep learning models are categorized into three separate models to evaluate performance under the circumstance with different feature sets. Model 1 is applied with syntax-based, sentiment-based, and grammatical-based features. The second model is applied with readability-based features, and the third model is applied with an amalgam of features used in the first two models. The comparison study has shown that the combined model performed well enough, with an accuracy of 86% compared to the first and second models. This experimental result is also compared with the LSTM model, which concludes that the Linguistic feature-based sequential neural network learning model accomplishes effective accuracy in a significantly short period of time.

To further fill the research gap of the lack combination of DL models for fake news detection, in [22], the author proposes a novel hybrid deep learning model that combines both RNN and CNN techniques that improve the accuracy of the classification task. Precisely, the model does binary classification between false information and legitimate news, and the result denotes either false or real. The hybrid CNN-RNN model utilizes CNN to capture local characteristics and make use of RNN to learn the long-term dependencies(sequential characteristics) of the dataset. The experiment was conducted on two publicly available pre-trained embedding datasets for performance evaluation. These two public datasets are the FA-KES: A Fake News Dataset around the Syrian War and the ISOT Fake News Dataset. For performance measures, the authors employed Accuracy, Precision, Recall, and F1-Score as measures and a paired t-test to confirm the result's statistical significance. As

a result, the proposed hybrid CNN-RNN model has the best performance, with accuracies of about 100% for the ISOT dataset and 60% for the KA-KES dataset. This work vastly improved the current baselines of the non-hybrid baseline methods. Usually, the neural network lacks generalization ability because of over-fitting on a specific training set or the specialization in certain tasks. According to the article [22], the novel approach becomes the first model that contributes to examining the model's generalization ability compared to the current majority of studies that commonly train and evaluate performance on the same dataset with stratified or random sampling.

Under the COVID-19 pandemic situation, COVID-19-related news massively emerged. As such, the spreading of fake news increased rapidly because of the increment in users' social media engagement because of the pandemic lockdown. The authors, in [23], proposed a Cross-SEAN model, which leverages the vast data without labels to learn the writing style of posts. This paper discussed relevant techniques for detecting fake news and content classification by semi-supervised models. It further uses a dataset of covid-19, CTF, and its examination of the Cross-SEAN model that could process largely unlabelled samples. For any news classification tasks, the Cross-SEAN model successfully avoids the major issues of lacking labeled data to train neural networks and the inefficiency of detecting the emerging fake news because the external knowledge is applied to this Cross-SEAN model that helps early detection and makes it become a generalized model. It also considers that external information from reliable references may be useful in classifying the data even if the model is not pre-trained on a specific task. The performance of Cross-SEAN achieves 0.95% that outperformed the other seven model baselines by at least 9%, namely, MTL [24], 3HAN [27], 1HAN (base version of 3HAN), 16HLT-HAN [26], CSI [28], dEFEND [29], and MixText [30]. The last contribution of this article is that it implements Chrome-SEAN to identify fake news for real-time and simple usage by users.

An alternate psychological discovery feasibly explained the rapid dissemination of COVID-19 fake news on social media. According to the confirmation bias theory, users tend to spread information that confirms his/her existing preferences. Users' online activities and behavior have great potential to predict users' preferences by fake news detection. The first step is to learn endogenous user preferences. The authors use the FakeNewsNet dataset [33] and utilize the Developer API to crawl past tweets from all users' posts in the FakeNewsNet dataset that retweeted the news. It further employs pre-trained BERT embeddings and pre-trained word2vec vectors to encode endogenous user preferences. Moreover, building a news propagation graph based on users' media platform engagement information. Since the previous work [35, 36, 37] proved that combining the user features with the news propagation graph is able to improve the performance of detecting fake news, in this way, this paper [32] proposes a framework called User Preference-aware Fake Detection (UPFD), which fuses endogenous and exogenous information using GNN. Hence, the UPFD predicts the credibility of news content on the social media platform. Six baselines used in this article are SAFE [38], CSI [28], BERT+MLP, word2vec+MLP, GNN-CL [35], GCNFN [37]. For performance evaluation in comparison of these baselines with the model UPFD, the observation is that UPFD exceeds the best baseline GCNFN by about 1%, which is ascribable to the advantage of modeling the users' endogenous preference.

Over the past few years, the automatic detection task that is based on crowd signals for digital media of news distribution (DMND) users has outstanding promising results in recent studies. However, it depends on the users' explicit opinions towards news which are challenging to attain. To decrease the usage of the explicit opinion, the article [25] proposes a better way to find misleading news in DMND through signals of the members of a crowd and from their reputations that does not require the explicit opinion of these users called Hybrid Crowd Signals (HCS). This approach used implicit user opinions, which are generated from the social media engagements regarding the dissemination of the analyzed news. HCS also uses the result of the classification of news as the machine's explicit opinion in such a manner that this model can be applied to any DMND using both implicit and explicit opinionators, even to the ones that are missing users' opinions, in contrast to Crowd Signals based approach. The model's performance is evaluated on five datasets, namely,

Gossip, PolitiFact, Gossip2, FakeNewsSet, and FakeBr. The experiments revealed a vast performance improvement of HCS with combined opinions. Besides that, it also demonstrates that the model is still competent without considering DMND users' explicit opinions.

### 3.2. Deep Learning Model

The neural network performs well on fake news detection tasks due to its complicated and large web of parameters. The inner workings of the network are always a mystery that is troubling the research community since deep learning has become popular. This section describes the model architecture of two deep learning techniques established by authors in [23, 32].

#### 3.2.1 Cross-SEAN

Here, we elaborate on various components of its architecture. The training data in [23] is comprised of labeled ($X_L$) and unlabelled ($X_U$) samples. The labeled sample consists of total $n_L$ data points and the unlabelled sample composed of an amount of $n_U$ Data points. Both labeled and unlabelled sample compose four sets, namely, external knowledge text, tweet features, user features, and tweet text, where each input sub-set are encoded respectively as the following:

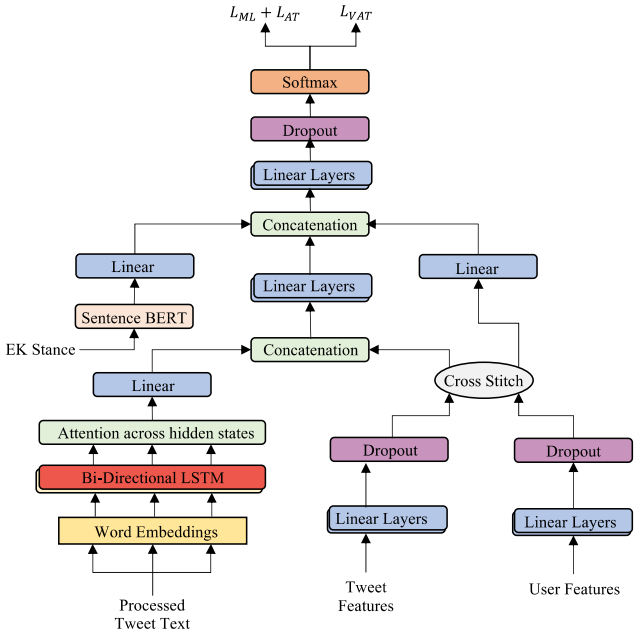



**Figure 2.** A schematic diagram of Cross-SEAN

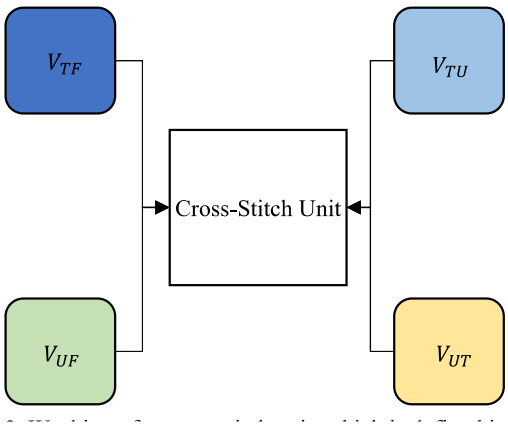



**Figure 3.** Working of a cross-stitch unit, which is defined in Eq.(9)

Encoding textual data: One-hot vector of vocabulary size V represents the tweet text of sequence with length N. The word embedding layer $E \in R^{V \times D}$ transforms the one-hot vector into a dense tensor $e \in R^{N \times D}$ which is $(e^1, e^2, \ldots, e^N)$. We use Bidirectional LSTM to encode the token vector,

where the forward LSTM releases a hidden state $h_{ft}$ each time. $h_{ft}$ is concatenated by a $h_{bt}$ of the backward LSTM to produce $h_t \in R^{2\times H}$, where H represents each LSTM's hidden width.

$$h_t = h_{ft} \oplus h_{bt},\ \forall t \in [1,\ N] \tag{5}$$

All layers output $f_k \in R^H$, and a net hidden vector that contains N vectors is merged with f using attention, which is defined as:

$$\mathrm{v} = \sum_{j=1}^{N} \mathrm{a_{ij}} h_j\ ;\ a_{ij} = Softmax(h_i \bullet f_j) \tag{6}$$

$$f = f_f \oplus f_b,\ f \in R^{2\times H} \tag{7}$$

$$h = h_1 \cdots \oplus \cdots h_N,\ h \in R^{N,(2\times H)} \tag{8}$$

We obtain the vector $v_{TT}$ to represent the encoded features of the tweet and find $e_{EK}$ of external knowledge matching to each input batch using the Sentence BERT. Finally, to obtain the $v_{EK}$ of the external information, pass the $e_{EK}$ vector through a linear layer.

Encoding tweet and user features: Firstly, we first pass $x_{TF}$ and $x_{UF}$ through distinct linear layers, which respectively interpolate them to the high dimensionality of feature vectors $v_{TF}$ and $v_{UF}$. Next, the cross-stitch units allow the model to choose the best mix of inputs from the features and portion across the multi-layers network. Through non-linear transformations, these units also provide a gradient flow route, where the transformation is defined as:

$$v_j' = \alpha_{ij} \bullet v_j + \beta_i,\ \forall i,\ j \in (1,\ K_T + K_U) \tag{9}$$

Where $a_{ij}$ and $\beta_i$ represent the weights of layers. The kernel of cross-stitch outputs $v_{TU}$ and $V_{UT}$.

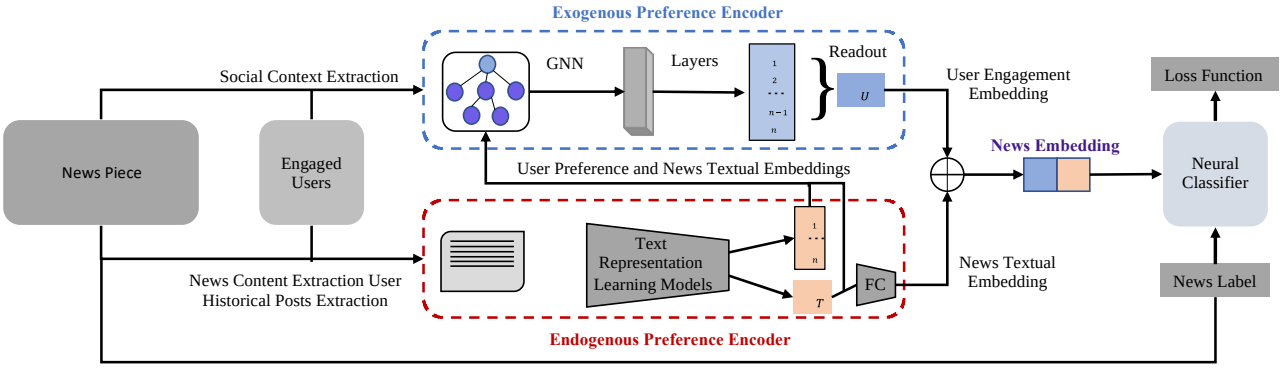



**Figure 4.** The UPFD model for fake news detection

Connected components in Network: The transformed feature vectors ($V_{TT}$ and $V_{TU}$) of the tweet's features and text is concatenated by equation $v_T = v_{TT} \oplus v_{TU}$ based on the inherent resemblance between tweet text and features vs. user features. The three feature vectors $v_T$, $v_{TT}$, $v_{TU}$ then go through distinct feed-forward linear layers, performing the affine transformation, and are concatenated to get the decoded vector v, which is eventually downscaled through NN and regularised by dropout layers. The likelihood is defined as:

$$\mathrm{p(y|x;\theta) = Softmax(v') = Softmax(\left|\left|v'_T\right|\right|\ \left|\left|v'_{U.T.}\right|\right|')} \tag{10}$$

Where $\mathrm{v'}$ represents the transformed vector, and $\theta$ defines the parameters at the present moment.

### 3.2.2 User Preference-aware Fake News Detection

This approach has three major components. Firstly, crawly the historical posts of its engaged users and its content on media platforms on social media to learn endogenous user preferences. To encode the news information and users' preferences, the authors adopt two types of text representation methodologies. Select 680k 300-dimensional vectors pre-trained by spaCy for word2vec and use

BERT embedding as a sequence with bert-as-a-service. The BERT framework encodes media and user data using a cased BERT-Large model.

To leverage exogenous user context, create the news propagation graph on Twitter, build a news article as $v_1$, and $\{v_2, \dots, v_n\}$ as the users that forwarded message $v_1$ in chronological order. Then, we can construct the news propagation graph as the following: (i): For all $v_i$ that retweets the same tweets later than one account in $\{v_2, \dots, v_n\}$, where $i \in N$, approximate the account's news spreads that has the latest timestamp to the account $v_i$. (ii): For all $v_i$ that subscribe zero users in the retweet sequences, as well as the original account, estimates the news spreads that have the maximum followers.

Next, combine the endogenous and exogenous information using a GNN encoder and apply readout functions overall node embeddings to get an embedding of the news propagation graph. Again by fusing process, the final news embedding is assembled of the textual and activities' embedding, which is put into a Multilayer Perceptron to predict the news' credibility.

## 4. Challenges

There are some aspects in our literature review of fake new detection to be considered challenges. Foremost, researchers lack generally accepted benchmark datasets due to the fact that verifying the news and social media posts is a cumbersome process for classification manually. These benchmark datasets are used to compare and evaluate the performance of state-of-the-art Fake News detection techniques. In this way, the acceptance of the dataset can be one of the fundamental factors that control approaches' performance. Such a dataset may not be adequate for building models that operate properly in real-world scenarios and identify the exogenous context. As a result, future studies in this area should concentrate on building large-scale datasets that fulfill the concerns of most real-world scenarios. Also, News Media Companies should release more data on fake news to deal with the bottleneck of lacking data, which further improves model performance during the experiments.

Even though many fake news detection methods achieve outstanding performance and are later implemented as real-world applications or browser extensions, the fake news is still spreading across social media since the user community has largely overlooked such tools. Firstly, most people do not ponder the validity of the news article or social media posts in an active manner and are more likely to publish and retweet what they believe. If we enable users with tools that validate the content as reliable on their social media accounts, it may result in a massive decrease in the spreading of rumors.

Finally, most research on contemporary fake news detection focused only on text data while the fake news is spreading in complicated ways, such as through text, images, and videos. It is recommended to examine the news article and the visual data, including images and videos, to capture the rumor through other formats, which help researchers build robust fake news detection systems.

## 5. Conclusion

Online media platforms have replaced the traditional media platforms with the increase in the use of the Internet in recent decades. The fake news disseminates through the Internet because of no control of higher authorities to guarantee what kind of news is accessible to users and is spread by users. As a result, some companies and celebrities use online media platforms for advertising their goods and maximizing their profits. Some users intentionally start to spread fake news in a different context to gain the advantage of their links and reputation. The problem of fake news affects millions of users' views which may cause a change of opinion of people. Hence, in this review paper, we analyzed a couple of research papers using various current fake news detection techniques and represented a taxonomy of these approaches. Then we explore the architecture as well as the metric to evaluate the model. Also, we should consider some possible future research directions for this task, such as focusing on the early detection of fake news to suppress the effects of the dissemination of

rumors. We genuinely believe that this review will provide a constructive opinion to help researchers gain concise insights into contemporary deep learning techniques and future trends.

## References

[1] Hermida, A. (2010). Twittering the news: The emergence of ambient journalism. Journalism practice, 4 (3), 297-308.

[2] Shiels, M. A. G. G. I. E. (2009). Twitter Responds on Iranian Role. BBC (17 June), http://news. bbc. co. uk/2/hi/8104318. stm.

[3] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of economic perspectives, 31 (2), 211-36.

[4] Addressing hoaxes and fake news. Meta. (2019, November 7). Retrieved June 5, 2022, from https://about.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/

[5] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter, 19 (1), 22-36.

[6] Zhang, H., Fan, Z., Zheng, J., & Liu, Q. (2012). An improving deception detection method in computer-mediated communication. Journal of Networks, 7 (11), 1811.

[7] Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B. J., Wong, K. F., & Cha, M. (2016). Detecting rumors from microblogs with recurrent neural networks.

[8] Zhao, Z., Resnick, P., & Mei, Q. (2015, May). Enquiring minds: Early detection of rumors in social media from enquiry posts. In Proceedings of the 24th international conference on world wide web (pp. 1395-1405).

[9] Devlin, J., Zbib, R., Huang, Z., Lamar, T., Schwartz, R., & Makhoul, J. (2014, June). Fast and robust neural network joint models for statistical machine translation. In proceedings of the 52nd annual meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 1370-1380).

[10] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 521, 436–444 (2015).

[11] Wang, W. Y. (2017). " liar, liar pants on fire": A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648.

[12] Singh, P., Srivastava, R., Rana, K. P. S., & Kumar, V. (2022). SEMI-FND: Stacked Ensemble Based Multimodal Inference For Faster Fake News Detection. arXiv preprint arXiv:2205.08159.

[13] Sepúlveda-Torres, R., Vicente, M., Saquete, E., Lloret, E., & Palomar, M. (2021, June). Exploring summarization to enhance headline stance detection. In International Conference on Applications of Natural Language to Information Systems (pp. 243-254). Springer, Cham.

[14] Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2018). Fakenewsnet: A data repository with news content, social context and spatialtemporal information for studying fake news on social media. arXiv preprint arXiv: 1809.01286.

[15] Nielsen, D. S., & McConville, R. (2022). MuMiN: A Large-Scale Multilingual Multimodal Fact-Checked Misinformation Social Network Dataset. arXiv preprint arXiv: 2202.11684.

[16] Boididou, C., Papadopoulos, S., Zampoglou, M., Apostolidis, L., Papadopoulou, O., & Kompatsiaris, Y. (2018). Detection and visualization of misleading content on Twitter. International Journal of Multimedia Information Retrieval, 7 (1), 71-86.

[17] Dou, Y., Shu, K., Xia, C., Yu, P. S., & Sun, L. (2021, July). User preference-aware fake news detection. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 2051-2055).

[18] Elhadad, M. K., Li, K. F., & Gebali, F. (2020, August). COVID-19-FAKES: a Twitter (Arabic/English) dataset for detecting misleading information on COVID-19. In International Conference on Intelligent Networking and Collaborative Systems (pp. 256-268). Springer, Cham

[19] Li, Y., Jiang, B., Shu, K., & Liu, H. (2020). MM-COVID: A multilingual and multimodal data repository for combating COVID-19 disinformation. arXiv preprint arXiv: 2011.04088.

[20] Ferreira, W., & Vlachos, A. (2016, June). Emergent: a novel dataset for stance classification. In Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies. ACL.

[21] Sahoo, S. R., & Gupta, B. B. (2021). Multiple features based approach for automatic fake news detection on social networks using deep learning. Applied Soft Computing, 100, 106983.

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

[23] Paka, W. S., Bansal, R., Kaushik, A., Sengupta, S., & Chakraborty, T. (2021). Cross-SEAN: A cross-stitch semi-supervised neural attention model for COVID-19 fake news detection. Applied Soft Computing, 107, 107393.

[24] Arora, U., Paka, W. S., & Chakraborty, T. (2019, August). Multitask learning for blackmarket tweet detection. In Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining (pp. 127-130).

[25] Freire, P. M. S., da Silva, F. R. M., & Goldschmidt, R. R. (2021). Fake news detection based on explicit and implicit signals of a hybrid crowd: An approach inspired in meta-learning. Expert Systems with Applications, 183, 115414.

[26] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016, June). Hierarchical attention networks for document classification. In Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies (pp. 1480-1489).

[27] Singhania, S., Fernandez, N., & Rao, S. (2017, November). 3han: A deep neural network for fake news detection. In International conference on neural information processing (pp. 572-581). Springer, Cham.

[28] Ruchansky, N., Seo, S., & Liu, Y. (2017, November). Csi: A hybrid deep model for fake news detection. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (pp. 797-806).

[29] Cui, L., Shu, K., Wang, S., Lee, D., & Liu, H. (2019, November). defend: A system for explainable fake news detection. In Proceedings of the 28th ACM international conference on information and knowledge management (pp. 2961-2964).

[30] Chen, J., Yang, Z., & Yang, D. (2020). Mixtext: Linguistically-informed interpolation of hidden space for semi-supervised text classification. arXiv preprint arXiv: 2004.12239.

[31] Choudhary, A., & Arora, A. (2021). Linguistic feature based learning model for fake news detection and classification. Expert Systems with Applications, 169, 114171.

[32] Dou, Y., Shu, K., Xia, C., Yu, P. S., & Sun, L. (2021, July). User preference-aware fake news detection. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 2051-2055).

[33] Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big data, 8 (3), 171-188.

[34] Twitter Developer. 2021. Twitter API. https://developer.twitter.com/.

[35] Han, Y., Karunasekera, S., & Leckie, C. (2020). Graph neural networks with continual learning for fake news detection from social media. arXiv preprint arXiv: 2007.03316.

[36] Lu, Y. J., & Li, C. T. (2020). GCAN: Graph-aware co-attention networks for explainable fake news detection on social media. arXiv preprint arXiv: 2004.11648.

[37] Monti, F., Frasca, F., Eynard, D., Mannion, D., & Bronstein, M. M. (2019). Fake news detection on social media using geometric deep learning. arXiv preprint arXiv: 1902.06673.

[38] Zhou, X., Wu, J., & Zafarani, R. (2020, May). $$\mathsf {SAFE} $$: Similarity-Aware Multi-modal Fake News Detection. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 354-367). Springer, Cham.