Fusion of Machine learning for Detection of Rumor and False Information in Social Network

Nehal Mostafa [*1], Ibrahim El-henawy [2], Ahmed Sleem [3]

[1] Department of Computer Science, Zagazig university, Sharqiyah, Egypt
[2] Faculty of Computers and Informatics, Zagazig University, Sharqiyah, Egypt
[3] Faculty of Computers and Informatics, Zagazig University, Sharqiyah, Egypt
Emails: nihal.nabil@zu.edu.eg; ielhenawy@zu.edu.eg; ahmedsleem8000@gmail.com

Abstract

In recent years, spreading social media platforms and mobile devices led to more social data, advertisements, political opinions, and celebrity news proliferating fake news. Fake news can cause harm to networks, communications, and users and cause trust issues toward government, healthcare, or social media platforms. This inspired many researchers to implement models to detect falsified information content. But there are still many issues that need to be discussed and explored. In our paper, we introduce categories of fake news detection methods and compare these methods. After that, the promising applications for false news detection are extensively discussed in terms of fake account detection, bot detection, bullying detection, and security and privacy of social media. After all, A thorough discussion of the potential of machine learning approaches for fake news detection and interventions in social networks along with the state-of-the-art challenges, opportunities, and future search prospects. This article seeks to aid the readers and researchers in explaining the motive and role of the different machine learning fusion paradigms to offer them a comprehensive realization of unexplored issues related to false information and other scenarios of social networks.

Keywords: fake news; social networks; machine learning; false information; networks; social media; fusion

1. Introduction

The spreading of mobile phones and social media platforms is considered a way for people to express their views and opinions and identify news. However, social media platform plays a vital role in transformation. But it has its drawback, for example, the Facebook scandal in the 2016 elections. Facebook sells the user's data to the Cambridge Analytica company responsible for the Trump campaign. This led to the wrong use of these data to spread fake information about the other candidate in elections. Also, during the Covid-19 pandemic in 2020, much fake news about the vaccine has been spread and shared in lots of language on social media. This creates trust issues regarding a vaccine. Francesco Rocca, president of the International Federation of Red Cross and Red Crescent Societies, says that fake news is the second pandemic.

Fake news detection becomes a vital area during dispersing of automated bots and the existence of security issues in social media. Fake news can be in the form of text, images, video, or audio. So, there is a need to perform fake

Doi: https://doi.org/10.54216/FPA.040105
Received: February 16, 2021 Accepted: Jun 15, 2021

news analysis on the different types of data using machine learning (ML) or deep learning (DL). ML and DL have achieved great success in different fields and areas of research. So many studies aim to identify false data using these methods. The spreading of fake accounts helps post spam messages, fake news, rumors, and illegal demand for money. Also, cyberbullying can harm people. All these issues must be detected and analyzed **[1]**.

Many different types of remotely sensed projects make use of the phrase "feature fusion," which has become more common in recent years. In spatial data classification and picture segmentation, texture characteristics have been combined. With the right approach, fusing many features may improve classification accuracy. When several features are normalized and combined into a single set of features, the user may concentrate on selecting the most useful features from this unified feature set. Feature fusion relies on the elimination of irrelevant information in favor of more meaningful data by focusing on characteristics that are associated with certain classes. In this article, we zero in on techniques for fusing features by using several prototypical cloud characteristics.

There are two main types of feature fusion techniques: those in which numerous feature vectors are fused by complicated vectors and then used to categorize the target features, and those in which the feature vectors are combined from tip to tail to build a new feature vector. The first approach is often employed for categorizing things.

This paper introduces a comprehensive survey of different categories of fake news detection methods. Also, we provide the most application for fake news detection, such as fake account detection, bots detection, cyberbullying detection, and security and privacy in social media. We discuss the most challenging area in fake news detection.

1.1. Motivations

Social networks are extremely rapid data engendering and broadcasting platforms. In contrast, millions and even billions of users are communicating on various web platforms generating a vast amount of data every second. However, in opposition to conventional sources of news (e.g., newspapers and news channels), the integrity of content spreading on social networks is questionable because of the liberation of free expression. In Recent Times, it could be observed that there has been a massive rise in the number of users available on social networks searching or publishing news and knowledge. The public content of social media is greatly affecting the users' selections of preferences. In this regard, the concept of "Fake news" has become pervasive after the "2016 U.S. presidential elections," in which it is thought that the falsified subjects distributed during the elections exert substantial consequences on the election marks. Motivated by this, this works review and analyzes different machine learning approaches employed to tackle the issues resulting from fake news by considering the present state of information pollution in terms of ecosystem, various platforms of data sharing and generation, intelligent analysis, and fact-checking tools. To deliver a thorough and systematic interpretation, this study concentrates on fake news detection in social networks through numerous research accomplishments regarding from machine learning perspectives, deliberating the architecture, improvements, challenges, and opportunities of machine learning for trustworthy social networks. This is commenced to provide some support in the descriptive characterization of the relationship between machine learning and reliable social networks by promoting efficient detection of rumors and fake news. More significantly, this study seeks to offer an important and comprehensive reference for follow-up researchers, the Artificial Intelligence community, machine learning engineers, and beginners.

1.2. Contributions

Unlike present survey studies, this work presents an inclusive coverage of discourse on the role of deep learning in detecting information pollution, rumors, and fake news through social networks.

- First, this work ultimately defined, categorized, and characterized the false information concepts in social networks to comprehend the underlying concept of false information and its main attributes. To the best of the authors' knowledge, this is the first attempt to address these notions collectively to afford the readers a complete understanding of the polluted information and its potential dangers and consequences.
- Second, this survey offers a comprehensive overview of the recent literature to clarify the role of machine learning algorithms and deep learning models in detecting fake news, rumors, and other forms of false online information. The discussion provides a detailed analysis of these approaches from

Doi: https://doi.org/10.54216/FPA.040105
Received: February 16, 2021 Accepted: Jun 15, 2021

different perspectives, including data modality, learning strategy, input representation, and detection stages. Also, the survey considers the intervention of fake news from a machine-learning viewpoint.

- Third, following the previous contributions, this work also provides and reviews the application areas related to the detection and intervention of fake news, including fake account detection, bot detection, bullying detection, and security and privacy of social media.
- Finally, the contemporary challenges related to identifying and detecting different categories of information pollution are highlighted, along with state-of-the-art and research gaps. This collectively provides great insights for determining opportunities for machine learning to struggle with the complicated question of deceitful content on social media, thereby improving the trustworthiness of social networks.

1.3. Paper Organization

The residue of this study is outlined as follows. Section 1 encapsulates the concepts and surveys related to fake news detection. The basics and foundations of fake news are discussed in Section 3. Section 4 discusses the research progress concerning Fake multimedia analysis. Section 5 argues the real-world applications of fake news detection. The challenges, opportunities, and future directions are reviewed in Section 6. Finally, section 7 concludes the study's contributions.

2. RELATED ADVANCES AND SURVEYS

In recent times, different survey studies have been introduced for fake news identification, detection, categorization, and prevention. The emphasis of each study varies from one to another in its target data, i.e., text-based, vision-based, voice-based, or mixed kinds of data. Some other surveys the different aspects of fake news related to the specific language. Thus, in the following, the most comprehensive and recent related survey studies are investigated and reviewed, then the main aspects that distinguish this study from the reviewed surveys are pointed out.

The authors provided an overview of the state-of-the-art technologies, approaches, benchmark data, and empirical configurations for social content and analysis of deceitful information socializing online. They also presented a taxonomy for categorizing false information and discussed the social impact of fake, enthusiasm for broadcasting false information, user discernment, as well as cutting-edge approaches for fact-checking. On the other hand, the authors surveyed and reviewed the methods of fake news detection based on four distinct perspectives: knowledge, style, propagation, and source. Knowledge-based approaches consider the detection of fake news by validating whether the knowledge in the news is dependable on real knowledge. The style-based approaches are concerned with the way in which fake news is written. The propagation-based approaches emphasize detecting fake news according to the way they propagate online. The source-based approaches emphasize detecting fake news by inspecting the trustworthiness of sources at different phases. Besides, the authors emphasize the reviewing detection of fake news in the context of Natural Language Processing by categorizing the conventional procedures for recognizing fake news, provoking the foremost dataset, and employing features to describe the fake news. They also reviewed the primary vectorization approaches schemes for transforming natural language information into scientifically practicable data. Moreover, the authors presented a systematic review of the literature that communicates the research work introduced to address Misleading information propagation over social networks with the main focus on the Spanish language. They also work to recognize awaiting duties for this society and challenges that necessitate synchronization among the prominent researchers on the topic. Furthermore, the authors performed a comprehensive overview of the multidisciplinary notion of social dishonesty and categories of online social deception attacks and their inimitable features contrary to other social outbreaks and cybercrimes. They also surveyed different defense techniques for inhibition, recognition, and reaction mitigation of online social deception attacks together with the relevant merits and demerits as well as the legitimate and moral worries related to that field.

This work differs from other related surveys in many standpoints indicated as follows:

- Firstly, this survey discusses and considers the definition of fake news in the context of present studies and real-world infrastructure, as well as their destructiveness to the public. This entails many concepts related to the concept of fake news rumor, false news, misinformation, clickbait, cherry-picking, satire news,

Doi: https://doi.org/10.54216/FPA.040105
Received: February 16, 2021 Accepted: Jun 15, 2021

disinformation, and deceptive news. Unlike previous survey studies, this work argues the challenges of defining fake news and presents an advanced, precise, and inclusive definition.

- Secondly, though recent works have investigated the significance of multidisciplinary research on fake news, this work outline a clear path towards that by performing a wide-ranging literature survey over different disciplines, recognizing an inclusive set of well-identified theories. The way these theories relate to fake news and its propagators is demonstrated along with the methods employing the theories both in the detection of fake news and intervention.
- Thirdly, existing surveys have typically restricted their scope to studying fake news detection from a particular research viewpoint. They mostly categorize the techniques of fake news detection by the kind of employed learning strategy or according to context information usage. In contrast, this study provides a comprehensive taxonomy for categorizing fake news detection in accordance with different perspectives.

3. DEFINITIONS AND FOUNDATIONS OF FAKE NEWS

3.1. Definitions

Until then, researchers have not agreed upon a standardized definition for fake news, even in journalism. An obvious and precise definition facilitates laying a consistent foundation for fake news analysis and assessing associated research efforts. However, finding a global definition for fake news is challenging take because of its overlapping with different categories of false information, as discussed in the next section. In this regard, according to the investigation of pertinent literature, the concept of fake news could be defined either at a narrow scale or broad scale. In the narrow-scale scenario, Fake news could be defined as deliberately falsified news circulated by a news retailer. This definition copes with the recent improvement in fake news detection, particularly after the 2016 U.S. presidential election. On the other hand, the broad-scale definition considers Fake news just false news. Whereas news generally incorporates articles, tweets, statements, speeches, and posts, among other forms of information associated with municipal figures and corporations. This definition assumes that either journalists or non-journalists can generate fake news. This definition provokes some societal matters, such that the concept of "fake news" has to be "about further than news" and "regarding the complete information ecosystem. Both scenarios necessitate the validity of fake news to be untrue or non-factual. As the objective is to offer a methodical definition of fake news; therefore, news untruths ought to be obtained by evaluating them against unbiased truths and not with personal perspectives. Thus, it is inappropriate to think about fake news as paragraphs that do not come to an agreement with persons' or groups' concerns or opinions, which is occasionally in what way the concept of fake news is employed by the general community or in political views. Such untruth could be appointed to the entire or portion of the news subject, or yet to true news, once following events have provided the fundamental fact outdated.

3.2. Categorization

Fake information is widely spread over social networks and can be presented in the form of tweets, posts, blogs, images, chats, stories, and defiant news. It is commonly known as information pollution and presents numerous arrangements that are not contradictory but simultaneously exhibit some heterogeneity that conveys them under an unambiguous group. Table 1 summarizes different categories and impacts of fraudulent content on the internet. Though each type of false information has some outstanding features, this study employs the name of these categories interchangeably in numerous sections to deliver a comprehensive collaboration of false information in social networks. The introduced taxonomy overview compares different overview categories of false information according to five aspects, including 1) description of each category, 2) authenticity, which represents the inclusion of any non-factual statement, and 3) intention, which characterizes whether the information aim to mislead or amuse the globe, 4) impact of each category on the public, and 5) the novelty of information.

Table 1: categorization and comparison between different kinds of false information

Category	Description	Authenticity	Intention	Impact	News
Clickbait	The deliberate use of misleading headlines to encourage visitors to	Undefined	Mislead	To earn advertising revenue, to trigger	Undefined

Doi: https://doi.org/10.54216/FPA.040105
Received: February 16, 2021 Accepted: Jun 15, 2021

	click on a particular webpage			phishing attacks	
Conspiracy theories	an explanation of an event that invokes a conspiracy by sinister and powerful actors, often political in motivation based entirely on prejudice or insufficient evidence	Undefined	Undefined	Extremely harmful to people and society	Undefined
Disinformation	Deliberately deceptive information with a predefined intention	Non-factual	Mislead	To promote a belief, idea, financial gain, or tarnish an opponent's image	Undefined
Fake News	False information spread under the guise of being authentic news is usually spread through news outlets or the internet to gain politically or financially, increase readership, and biased public opinion	Non-factual	Undefined	to damage an agency, entity, or person or gain financial/political profit	Yes
Misinformation	Circulating information that becomes false inadvertently because of an honest mistake, carelessness, or cognitive bias	Non-factual	Undefined	Less harmful but wrong interpretations of facts can lead to big damage	Undefined
Opinion Spam	Fake or intentionally biased reviews or comments about products and services	Non-factual	Mislead	untruthful customer opinion	Yes
propaganda	Unfairly prejudiced and deceptive information spreads in targeted communities according to a predefined strategy to promote a particular viewpoint or political agenda.	Commonly factual	Mislead	Political/financial profit	Undefined
Rumor	An unverified piece of information that is not necessarily false; may also turn out to be true.	Undefined	Undefined	Uncertainty and confusion about facts	Undefined
Satire/parody	Articles that mainly include humor and irony have no damaging aim but can deceive. The Onion and Satire Wire are sources of satirical news commentaries.	Non-factual	Entertain	The motive is fun but sometimes exerts adverse effects also	Yes

3.3. Fundamental Theories

Essential humanoid intellect and conduct theories have evolved throughout many disciplines, i.e., societal sciences, medicine, and finances, delivering indispensable intuitions for fake news analysis. These theories might provide new prospects for categorical and numerical fake news analysis in the era of big and could also enable the development of well-validated and reasonable models for detecting and intervening in fake news that is dominating social networks every day. According to a deep investigation of the literature in different disciplines, it is recognized that there. are common theories that could be hypothetically exploited for fake news analysis.

First, news-related theories have disclosed the conceivable features of the content of fake news in opposition to true news content. For example, theories have indicated that fake news hypothetically varies from true ones in different ways, such as writing style and quality of words, the number of words, and well-stated emotions. It ought to be remarked that these theories, created with forensic psychology, focus on misleading statements or disinformation

Doi: https://doi.org/10.54216/FPA.040105
Received: February 16, 2021 Accepted: Jun 15, 2021

instead of fake news, even if these are comparable conceptions as previously discussed. Therefore, this opens a promising opportunity for the research community to confirm whether these characteristics are statistically distinct from one category of false information to another, i.e., fake news, truth, and disinformation, specifically making use of fake social news. In contrast, these discovered discriminatory characteristics could be employed to instinctively distinguish fake news employing its writing design, whereas standard studies often employ supervised machine learning to accomplish this task.

Second, the user-related theories scrutinize the attributes of clients engaged in fake news events, such as tweeting, redirecting, posting, reacting, and commenting. Unlike fake reviews, fake news may attract nasty and typical users. In contrast, malicious users deliberately propagate fake news and are powered by the required profits. Not Many typical users (i.e., typical susceptible users) could habitually and accidentally propagate fake news without realizing the deception. Such susceptibility mentally originates from either self-impact or social impacts, where theories have been appropriately classified and detailed in Figure. Unambiguously, as designated by the normative influence theory, bandwagon consequence, an availability cascade, and social identity theory, to be fond of and agreed upon by the community, typical users are urged to participate in the action of fake news under situations where several users undergo that. One's confidence in fake news and his/her unintended dissemination could be endorsed too in case of being subjected more to fake news, which habitually happens because of the echo chamber influence over the social network.

This confidence in fake news could be constructed once they demonstrate one's established opinions, views, or propositions, which are habitually supposed to go beyond that of others and have a tendency to be inadequately amended once new contradicting testimony is offered. In this regard, schemes for fake news intervention according to the user's viewpoint ought to be carefully intended for users with various heights of integrity or objectives, despite the fact they could all participate in similar fake news actions. As an example, it is acceptable to interfere with the propagation of fake news by punishing malevolent users only. Rather than, learning and subjective suggestions of genuine news and disproven fake counterparts may be beneficial for typical. users. These suggestions have to not only accommodate the subjects that the users would like to read but must also describe subjects to which users are most naïve.

3.4. Multimedia type in fake news

Text: linguistics aims to analyze the text depending on its phonetics, phonology, morphology, syntax, semantics, and pragmatics.

Image/video: content in social media can be attached with fake images and videos, which can be detected using Machine Learning (ML) or deep learning (DL) techniques.

Audio: This media type is spread via WhatsApp, Clubhouse, and many podcast apps.

3.5. Multimedia Type of model approaches

As a result of the above, according to Zeng et al. **[2],** most studies aim to detect text misinformation (unimodal) or detect many media misinformation (multimodal).

Unimodal misinformation detection: Most studies concentrate on identifying rumors by evaluating only text contents using ML or DL methods.

Multimodal misinformation detection: The increase in mobile device usage leads to more unitization of social media, which increases text posts combined with other multimedia types. Many studies perform multimodal feature integration during inadequate text-only analysis **[2]**.

4. FALSE INFORMATION ANALYSIS

Recently, many research efforts have investigated and applied machine learning algorithms to detect fake information, and they achieved a great performance compared to traditional false information techniques. In this regard, this section reviews the state-of-the-art ML and DL literature for fake news detection. For convenience, the machine learning approaches are taxonomized according to five different aspects, as shown in Figure 1. This includes learning strategy, modality, input representation, type of data, and detection stage. In the following subsection, we provide an in-depth discussion and comparison of different studies belonging to different categories of approaches.

Doi: https://doi.org/10.54216/FPA.040105
Received: February 16, 2021 Accepted: Jun 15, 2021

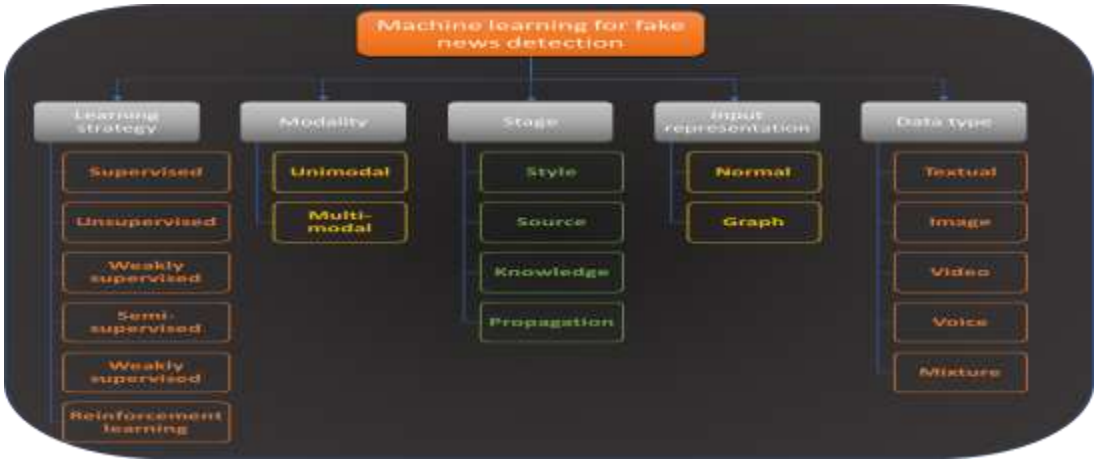

Figure 1: taxonomy of machine learning approaches for fake news detection in social networks.

A) Unimodal analysis

The unimodal analysis uses natural language processing (NLP) techniques to convert unstructured data to structured data that can be analyzed using machine learning (ML) or Deep learning (DL) algorithms to extract meaningful information from it. Fake text analysis passes into different stages, such as pre-processing stage, feature extraction, and detection techniques. Each one of these stages has several proesses that are performed on the text **[3]**.

Pre-processing stage: This stage involves several processes to convert unstructured data to structured data. One of these processes is tokenization which breaks down the text into tokens. This process can be performed on words, text paragraphs, articles, or chapters. The main aim of tokenization is to convert text into meaningful constituents by removing whitespaces or converting upper cases to lower cases. But this process differs from one language to another. For example, in the Chinese language, there is no white space, and in the Arabic language, many elements are connected to a word. This process is essential in fake text analysis, so it is hard to deal with text data without tokenization **[3]**.

Another process part of speech (POS) tagging aims to assign each token to a corpus (text) and identify if the word is a noun, verb, adjective, etc. This process produces a form of (word, tag). After that, the stemming process reduces terms to their roots, while lemmatization reduces forms to the base form **[3]**.

Feature extraction: Most social media data is considered large-scale or big data. In the previous stage, the data is prepared in a structured form. So, the main aim of this stage is to retrieve metadata that defines the document in vector form. One of these methods is the term frequency-inverse document frequency (TF-IDF) which indicates the terms that appear in a document with a high frequency (T.F.) but in the total collection of which the document is part at a low frequency (IDF). So, it indicates the importance of the term within the document, which is computed as follows:

$$tf_{t,d} = \frac{n_{t.d}}{|d|} \quad (1)$$

$$idf_t = \log_{10} \frac{N}{df_t} \quad (2)$$

$$w_{t,d} = td_{t,d} * idf_t \quad (3)$$

Where $tf_{t,d}$ is the term frequency of term t within document d, $n_{t.d}$ is the number of term t in d. $|d|$ is the size of a document? N is the number of documents in the collection. To evaluate the relevancy of a query to a document, the vector space model (VSM) is using **[3]**.

Doi: https://doi.org/10.54216/FPA.040105
Received: February 16, 2021 Accepted: Jun 15, 2021

Detection techniques: In this stage, the document is ready for classification and clustering tasks to define the meaningful information from the document. Suppose the processing will be performed on such a reasonable amount, so it is perfect for implementing ML algorithms such as Support Vector Machine (SVM), Naïve Baye, and K-nearest neighbor (KNN). But if the processing will be done on a large amount of data, it will be reasonable to use deep learning (DL) algorithms such as Convolution neural network (CNN) or recurrent neural network (RNN). The data is classified as big data if it matches the criteria of 5v's of big data (volume, variety, veracity, value, and velocity) [1].

Also, some studies implement a fake detection model, such as BERT, ROBERTA, ELECTRA, DistilBERT, and ELMov, XLNet **[1]**. Most of these tools are an improvement for the BERT model, which is summarized in **Table 2**. In the following, we will discuss most studies in fake text analysis (unimodal detection) for both web and social media fake news using ML or DL algorithms. Also, we summarized these procedures in **Figure 2**.

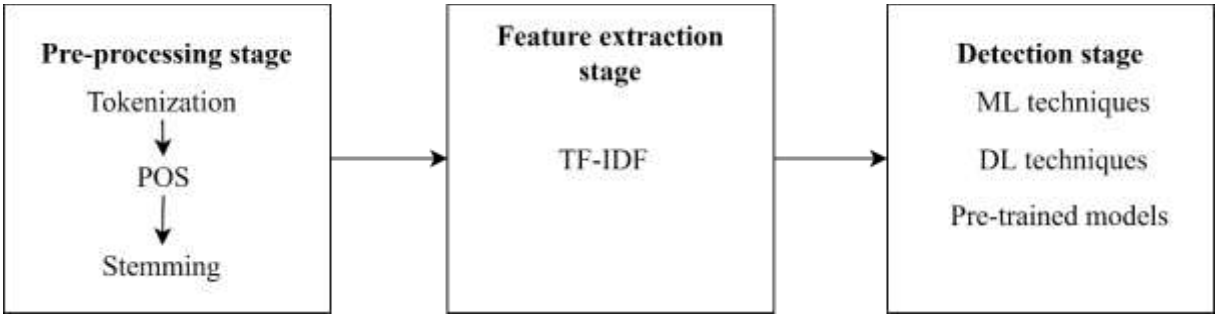

Figure 2: Unimodal fake news analysis stages

Table 2: Summarization of fake detection language models

Method	Characteristics
BERT (Bidirectional Encoder Representations from Transformers)	Pre-trained bidirectional model to learn the unlabeled text
Roberta (Robustly optimized BERT approach**)**	The bigger size mini-batches, No NSP loss and dynamic mask pattern change
ELECTRA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately**)**	Pre-trained using another former language, very effective in small data
DistilBERT	Smaller and faster, sufficient for production level
ELMo (Embeddings from Language Models**)**	Trained on big corpus using Deep bidirectional language
XLNet	Bigger data and more computational skills than BERT, bidirectional transformer with permutation-based modeling

Fake text analysis studies aim to distinguish fake and legit data. Some studies extracted sentiment features, such as Rubin et al. **[4]** predicting the satire news using 5 features. Another ML method aims to analyze n-gram features. This approach compares 6 text classifiers, and the results show superior accuracy for TF-IDF and SVM **[5]**. Also, a contribution by Bourgonje et al. **[6]** proposed a method to detect clickbait. Clickbait is a link in web pages that encourages visitors to click on it. This approach uses two-step logistic regression (L.R.) for clickbait detection depending on lemmatization for n-gram matching. A contribution using a Naive Bayes (N.B.) classifier exploits the similarity between false news and spam emails **[7]**. For the empath feature, an approach by Fast et al. **[8]** aims to create many lexical classifications using the empath tool, which integrates the NLP and Linguistic Inquiry Word Count (LIWC), which provides greater text analysis performance.

Doi: https://doi.org/10.54216/FPA.040105
Received: February 16, 2021 Accepted: Jun 15, 2021

In DL fake text analysis methods, Khan et al. **[1]** provide a comparison between many DL methods hybrid with ML methods. This study used combined corpus, fake or real news, and liar datasets to show its results. In ML approaches, N.B. shows higher results than other algorithms on the combined corpus dataset. In DL approaches, CNN shows a superior result to other methods on the Liar dataset, while LSTM and Bi-LSTM showed an overfitting problem. A new proposed hybrid Conv-HAN achieved higher accuracy on the Liar dataset. A use of CNN to classify unlabeled data by Liu et al. **[9]** proposed a method that inputs the sensitive feedback statutes through the process.

For advanced pre-trained models, Khan et al. **[1]** provide a comparison between many pre-trained models such as BERT, ROBERTA, ELECTRA, DistilBERT, ELMo on Liar, Fake or real news, and combined corpus datasets. The results show that models can deal with small datasets without overfitting problems. Also, these models have a superior F-score over ML or DL models. Also, it is observed that Elmo has a lower accuracy than other models. A hybrid pre-trained and DL model **[10]** between Bert and CNN to detect fake articles collected during U.S. presidential election 2016. These discussed studies are shown in **Table 3.**

Table 3: Summarization of most studies in unimodal analysis detection

Authors	**ML, DL, pre-trained model**	**Method**	**Web content (W) or social content (S)**	**Dataset**	**Evaluation metric**
Rubin et al. **[4]**	ML	sentiment methods	W	Collected	Precision = 90%, Recall = 84%, F score = 87%
Ahmed et al. **[5]**	ML	TF-IDF, SVM	W	Collected	Accuracy = 92%
Bourgonje et al. **[6]**	ML	LR	W	First Fake News Challenge (FNC1)	Accuracy = 89.59%
Granik & Mesyura **[7]**	ML	Naïve Bayes	Combination	Collected by BuzzFeed News	Accuracy= 74%
Fast et al. **[8]**	ML	Empath tool	Combination	Collected	Empath with LIWC correlation= 0.906
Khan et al. [1]	ML	NB	W	Combined corpus	Accuracy= 93%
Khan et al. [1]	DL	Conv-HAN	W	Liar	Accuracy= 0.59
Khan et al. [1]	Pre-trained model	BERT	W	combined corpus	Accuracy= 95%
Khan et al. [1]	Pre-trained model	RoBERTa	W	combined corpus	Accuracy=96%
Khan et al. [1]	Pre-trained model	ELECTRA	W	combined corpus	Accuracy=95%
Khan et al. [1]	Pre-trained model	DistilBERT	W	combined corpus	Accuracy=93%
Khan et al. [1]	Pre-trained model	ELMo	W	combined corpus	Accuracy=91%
Kaliyar et al. **[10]**	DL and Pretrained model	FakeBERT	W	Real-world fake news dataset	Accuracy= 98.90%
Liu et al. **[9]**	DL	Fned	S	Twitter and Weibo	Accuracy=90%

Doi: https://doi.org/10.54216/FPA.040105
Received: February 16, 2021 Accepted: Jun 15, 2021

B) Multimodal analysis

The spreading of mobile devices helps to enrich web and social content with information and the latest news. But it also allows the diffusion of fake news and information in any type of data, whether in text, image, video, or audio data. So the research community introduced a model that can do multi-analysis for many types of data named multimodal analysis. In this section, we provide a review of most studies on multimodal analysis.

Zeng et al. **[2]** introduce the FND-SCTI model that can detect false news by finding the semantic correlations between text and images. First, the image representation uses the VGG DL model and then uses a multi-variant autoencoder (VAE) to learn text and image representation. Finally, perform an eigenvector to detect the false news. Another contribution is using a generative adversarial network (GAN) that detects the new event using event variant features. This approach does not consider images if the dataset does not contain images **[11]**. Khattar et al. **[12]** use VAE to learn represented features of text and images. This approach improves the text representation using visual features. A new approach, CARMN **[13]**, is cross-modal attention residual, multichannel convolution neural network. We summarize all these studies in **Table 4**.

Table 4: Summarization of most studies in multimodal analysis detection.

Authors	ML, DL, pre-trained model	Method	Multimedia type	Dataset	Evaluation metric
Zeng et al. **[2]**	DL	FND-SCTI	Text and image	Twitter, Weibo	Accuracy=83.9%
Wang et al. **[11]**	DL	EANN	Text and image	Twitter and Weibo	Accuracy=86.6% (Weibo)
Khattar et al. **[12]**	DL	Mvae	Text and image	Twitter and Weibo	Accuracy =82.4% (Weibo)
Song et al. **[13]**	DL	CARMN	Text and image	Twitter and Weibo	Accuracy=92.2% (Weibo)

5. APPLICATION ON FAKE NEWS DETECTION

5.1. Fake accounts detection

In the last few years, a huge number of new accounts have been registered on many social media such as Facebook, Twitter, Instagram, etc. But it found that most of these accounts are fake and can harm the network or people on social media. Fake accounts are considered the basis for fake news and misinformation. So, it is essential to filter fake from real accounts. Many studies identify the features of fake accounts. In this section, we will discuss fake account detection contributions. The fake account features can be classified as account features or textual features **[14].** Account features can be a username, profile picture, number of followers, number of likes, and location. Textual features can be sender, mentions, hashtags, links, and number of replies **[14]**.

Khaled et al. **[15]** use a combination of SVM and N.N. to detect fake accounts and achieve high accuracy using Spearman's rank-order correlation technique. Another study introduces a method that detects relaxed functional dependencies (RFD) to identify fake accounts **[16]**. Further account-based feature approach **[17]** aims to evaluate profile similarity communication matching to identify the replicated accounts. The username feature-based utilized by Rahman et al. **[18]** detects fake Twitter accounts using fake project data. This approach implemented many classifiers, but random forest (R.F.) is shown to have the highest accuracy. One drawback of this method was the small nonrealistic number of accounts used in data, leading to different results in real time. Another use for the R.F. classifier is to detect fake accounts based on the emotion feature. The hate, ugly, etc., define the fake account **[19].**

In textual feature-based, Swe and Myo **[20]** introduced a model to identify fake accounts depending on a blacklist which was implemented using the topic keyword. This approach does not need profile or network-based features, which help to decrease the time and cost. Another approach by Khan et al. **[21]** separated spammers and bloggers by using the Hyperlink-Induced Topic Search (HITS) method. Further contribution

Doi: https://doi.org/10.54216/FPA.040105
Received: February 16, 2021 Accepted: Jun 15, 2021

used seven features to find a compromised profile. This approach creates a user history to determine fake or no accounts **[22]**.

In the hybrid feature, some studies detect fake accounts on Twitter by using a combination of graphics and content-based features and many classifiers. The results showed superior results for the R.F. classifier **[23]**. Another study by Aswani et al. **[24]** proposed an approach to detect twitter spammers by utilizing the firefly algorithm. Further contribution can identify both spam messages and fake accounts by extracting eighteen features and using an R.F. classifier **[25]**. We summarize all these studies in **Table 5**.

Table 5: Summarization of most studies in fake account detection.

Author	Dataset	Account-base or textual-base	Feature	Result
[15]	Twitter, MIB	Account	Profile photo	Accuracy= 98%
[16]	Twitter	Account	Biography	-
[17]	Facebook	Account	Profile similarity	Accuracy=93.87%
[18]	Fake project	Account	Username	Accuracy=99.2%
[19]	Facebook	Account	Emotions	Accuracy=90.91%
[20]	Twitter and honey pot	Textual	Number of Tweets	F-measure=0.954
[21]	Twitter	Textual	Hashtags	Precision = 0.7
[22]	Twitter	Textual	Sent Date	Precision =99.01
[23]	Twitter	Hybrid	No. tweets, followers, likes	F-measure=0.91
[24]	Twitter	Hybrid	No. tweets, following count, followers, retweets, listed count, mentions, hashtags, and links.	Accuracy=97.98%
[25]	Twitter	Hybrid	Biography, Birthdate, location, No. tweets, following count, Retweets, Mentions, Hashtags, and links.	Precision=0.933

5.2. Bot detection

Bots is a software robot that is programmed to automatically perform a certain task. Bots can be used in social media to operate a platform or post recent articles or news. But on the other side, malicious bots can generate Sybil bots, spam bots, social bots, and cyborgs bots **[26]**.

Sybil's bots: This bot can generate fake accounts that can spread fake information or malware over the network. A complicated authentication mechanism will prevent Sybil's attacks.

Spambots: The bot is able to send spam messages, links, or any junk data. Sending a bulk of fake data can congest the network.

Social bots: The bots collect sensitive data from people using fake webpage which act like the original ones.

Cyborg bots: Any type of bot that a human can control.

Many studies implement methods to prevent these bots. These studies are classified as graph, ML, crowdsourcing, and anomaly **[26]**. The graph is a set of vertices and sets of edges which sufficient in the social network representation. So many contributions aim to detect malicious bots using graph methods. Cornelissen et al. **[27]** integrated the ML methods with network measures. The results on the Twitter dataset showed unacceptable accuracy. Another study aims to detect political bots using founded strongly connected users by presenting two models post-to-post and user-to-user **[28]**. Moreover, Abu-El-Rub and Mueen **[29]** introduced a Bootcamp to identify the campaigns of bots using graph methods for topographical modeling to perform bot clustering.

ML methods achieve high performance in many fields. ML learning can be supervised learning (S.L.) or unsupervised learning (USL), or semi-supervised learning (SSL). The SL is the most used method in ML, which uses labeled input and output. An example used S.L. methods to detect malicious bots, Khaled et al.

Doi: https://doi.org/10.54216/FPA.040105
Received: February 16, 2021 Accepted: Jun 15, 2021

[15] introduced a hybrid SVN-NN to detect Twitter bots. Moreover, Daouadi et al. **[30]** provided a feature on fake accounts depending on interactions with other accounts and users. This approach uses a deep forest algorithm to detect bots on Twitter.

In USL, learning is without any labeled output. An example of detecting bots using USL, Chew **[31]** aims to identify automated bots to detect fake accounts by finding patterns of similarities. This approach depends on finding emergent patterns for groups of accounts. Another contribution by Chen and Subramanian **[32]** introduced an approach to finding spam bot campaigns on Twitter. By founding the latest URL tweets on real-time streaming, so the approach can find the account that spread the same texts and define it as spam.

SSL learning is a combination of SSL and USL learning which uses incompletely labeled data. An example of SSL-detecting bots by Shi et al. **[33]** introduced an approach to detect social bots by using clickstream sequences and SSL for clustering. Another contribution based on the homophily property of social network graphs, a SocialBotHunter model, is introduced to control users' behavior and interaction on social media **[34]**.

Crowdsourcing is behavior on social media that asks users to perform manually. Such as online surveys to enchant the service. Crowdsourcing can detect Sybils and cyborgs, as Alarifi et al. **[35]** included ten crowd workers to manually detect real Sybils, and Cyborg accounts on Twitter. Moreover, Cresci et al. **[36]** aim to evaluate the accuracy of crowd worker detection. The results show high detection accuracy in traditional Spambots and genuine accounts and low detection for social Spambots.

Anomaly detection aims to find odd behavior in users or groups. Such as selecting random English tweets of six million users found that the group of accounts are bots. These accounts shared odd tweets such as random quotes**[37]**. Also, Lee et al. **[38]** based on the strange behavior of spam profiles interacting with junk profiles to increase their contacts' network. Utilizing the honeynet strategy, discover the junk information on Twitter. The authors created 60 honeypots mimicking junk content creators. If a real account follows the honeypots, so is classified as a junk account. The same procedure was done on the Arabic dataset **[39]**. We conclude this study in **Table 6**.

Table 6: Summarization of most studies in BoT detection.

Author	Method	Dataset	Result
[27]	Graph-base	Twitter	Accuracy=70%
[28]	Graph-based	Reddit	-
[29]	Graph-based and ML	Twitter	Accuracy = 87.5%
[15]	ML(SL)	Twitter	Accuracy= 98.3 %
[30]	ML (SL)	Twitter	Accuracy=97.55%
[31]	ML (USL)	Twitter	-
[32]	ML (USL)	Twitter	-
[33]	ML (SSL)	Twitter	F-score=95.2%
[34]	ML (SSL)	Twitter	Accuracy=99.5
[35]	Crowdsourcing	Twitter	F-measure=91%
[36]	Crowdsourcing	Twitter	Traditional Spam bots =91.36% Genuine accounts= 92.01% Social Spam bots=23.55%
[37]	Anomaly detection	Twitter	-
[38]	Anomaly detection	Twitter	-
[39]	Anomaly detection	Twitter	F-measure=76.55%

5.3 Data Fusion

In the next section, we give a thorough evaluation of the existing ML-based methods for detecting fake news. These techniques may be used to analyze many forms of digital material. Here we provide an overview of the techniques used in I message and Natural Language Processing (NLP) analysis,

Doi: https://doi.org/10.54216/FPA.040105
Received: February 16, 2021 Accepted: Jun 15, 2021

Separating Meaning from Text

From a purely intuitive perspective, natural language processing (NLP) seems to be the best method for automatically detecting bogus news. Extracting characteristics from the content of the examined article is the fundamental information source required to construct a trustworthy pattern recognition system, notwithstanding the importance of the sociocultural settings of the message delivered through electronic media. Several overarching themes may be picked out of the works done in this field. Analyzing text without linguistic structure is the easiest conceptually; this is most usually done using a backpack (originally suggested in 1954 by Harris) or N-grams. However, cognitive and metacognitive variables, semantic research, and other similar methods are also widely employed.

NLP-based data representation

Saquete et al. note that each of the subprojects differentiated within the area of fake news detection relies on the resources provided by NLP, and this is an important point to make. Bag-of-words, a simple approach that just counts how often certain words appear in the text, is the foundation of many advanced techniques. N-grams, which tokenize whole strings rather than single words, are a little more advanced extension of this principle. Bag-of-words may be defined as unigrams if N is equal to one, which is within the scope of the concept. Also, the total number of N-grams in a text is strongly reliant on its length, therefore this metric has to be normalized either to the document (Term frequency: TF) or to the collection of documents utilized in the learning process before pattern recognition systems can be built (TF–IDF).

These ideas, which are both straightforward and relatively old, are nevertheless put to good use in the modern world to combat the issue of identifying fake news. Hassan et al .'s work is a good illustration of this type of application; they tested five different classification models (Lin-SVM, RF, LR, NB) on the task of classifying disinformation on Twitter and compared their performance to that of attribute extraction techniques based on term frequency and term frequency-inverse document frequency (TF-IDF) using N-grams of varying lengths. The PHEME dataset was used to demonstrate the efficiency of extracting both short and long word sequences at the same time. Most investigations begin with the same steps, which leads to further recommendations for things to think about, such as a thorough examination of base classifiers or the use of ensemble techniques.

Other exciting developments in this area of study include the rejection of unanswered questions, emojis, and numerous exclamation points during the preprocessing stage, as well as the recognition of unique tokens such as text components, repeatedly denied statements, and swear words. Deep networks are an exciting area of AI research, similar to their success in many other domains of use. Because of their widespread use as an option to traditional models, they are contributing to this discussion as well.

In order to make the most of the connections among the numerous actors discussed in a news item, our method entails the following stages.

Data mining for metadata Step one is to gather the accessible metadata, including user feedback, entity information, and any other relevant data, from third-party sources.

Claiming Connections As a second step, we'll compile the broad relationships that can be identified in a KG, and (ii) the specific ones that can be calculated from the available texts.

Word embedding The next stage is the modification of various brain models (e.g., via including a layer of word embedding) for better false news identification.

Features provided in earlier phases may also be external, but those included in the last stage will be inside exclusively. A schematic of the complete operation is shown in Figure. 3.

The current data modeling that resulted in the extra descriptors was motivated by the belief that, by including derived entities and distinguishing clearly between immediate and indirect speaking, we can lay the groundwork for more nuanced discussion that may pin down the speaker's prior interactions with the problem in question (or subject) and with all concerned parties in the corresponding issue. In the future, such a study might be expanded to provide even more nuanced information about a speaker, such as whether or not an individual identified as a

Doi: https://doi.org/10.54216/FPA.040105
Received: February 16, 2021 Accepted: Jun 15, 2021

politician really supports the policies of their party. Thus, it allows the graph to be used as a window into the inner workings of different statements.

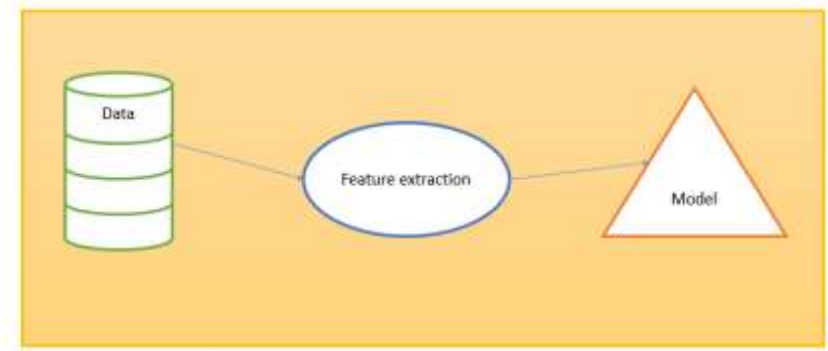

Figure 3: The neural network with the semantic features.

We use the Liar dataset. The Liar is mined for short political-themed texts that are then categorized according to their level of veracity; the Liar data sets include these texts and also include credit histories that monitor the veracity of the speaker's remarks. There are three distinct sections of the data collection (training, testing, and validation). Table 7 shows the description of the dataset.

Table 7: The training, testing, and validation of the dataset.

	Train	Validation	Test
Split Dataset	8213	851	890

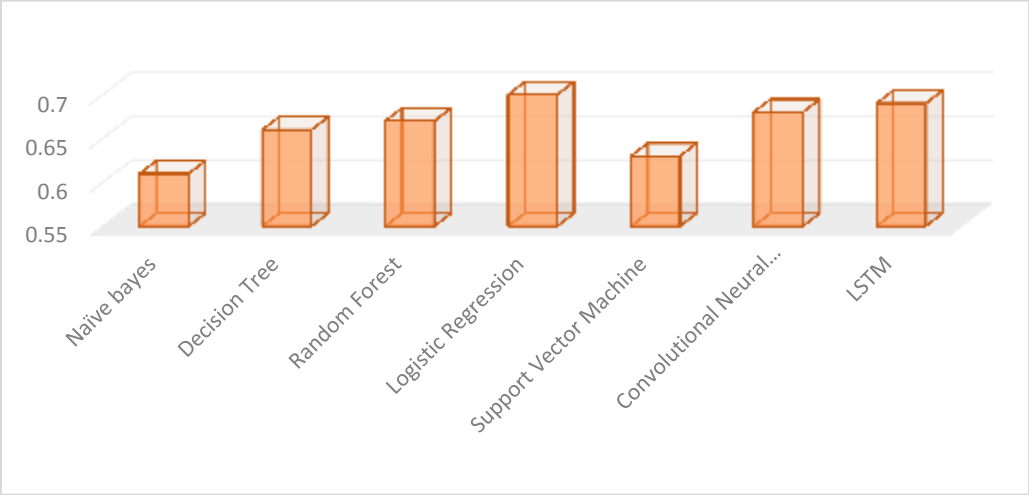

Figure 4: The accuracy of the dataset.

To begin, we put scikit-learn-created "traditional" models through their paces. Models that use relational characteristics often improve by a few percentage points over those that just use the original features of the data source. However, even the highest scores are subpar. The ratings for logistic regression are the best algorithm. In one instance (the random forest classifier), we find that using the additional relational information does not outperform using simply the original text. The naïve Bayes turns out to be the worst "traditional" ML classifier.

Phase 2 involves putting several DL models through their paces. The DL models use the hot encoding of the different classifiers and are developed using Keras and TensorFlow. The accuracy of the Adam optimizer is used as the reported evaluator measure for DL models.

Doi: https://doi.org/10.54216/FPA.040105
Received: February 16, 2021 Accepted: Jun 15, 2021

This research employs a variety of DL classifiers, all of which utilize the Adam optimizer and the categorical cross-entropy loss function.

The models are oversimplified in CNN. It's made up of the standard hidden layer, a Convolution1D that "learns" to filter groupings of words, an embedding layer with a dropout of 0.25, and a MaxPool layer for optimizing the network's performance (dense, dropout set to 0.25 and relu activation function). A single softmax-squished output layer receives the projection.

Dimensionally, BasicLSTM is a 400-by-400 LSTM with a MaxPool layer, a spatial dropout of 0.2, and softmax activation in its dense layers. The batch size is 128, the epochs are 50, and the learning rate is 0.00001. Figure 7 shows the accuracy of the previous models.

The Buzzfeed dataset

In reality, there were three separate datasets combined to form the one used: I Buzzfeed, which included Facebook data related to the US Presidential Election; (ii) Political news, which included news from reputable outlets like The Guardian and BBC as well as less reputable ones like Ending the Fed and Infowars as well as satirical outlets like The Onion and SatireWire; and (iii) Burfoot and Baldwin, which included mostly real news stories from 2009. The whole dataset isn't balanced, unfortunately. There are 4111 factual stories, 110 fictional ones, and 308 humorous ones.

The CREDBANK data set was first presented. Extending back to October 2015, this massive crowdsourced dataset contains over 60 million tweets across 96 days. Over a thousand different news events are covered, and 30 Amazon Mechanical Turk editors vet each tweet for veracity.

Periodically updated, the FakeNewsNet repository was suggested by the authors. This dataset is a compilation of Snopes and BuzzFeed content that has been republished and shared on Twitter, including the original article's metadata (source, body, multimedia) and the original poster's and recipient's social profile information (user profile, followers/followee).

The ISOT Fake News Dataset is the next to be discussed. It has a healthy distribution of both fake and real news stories, totaling approximately 12,600. The data set was assembled from primary sources; accurate information was culled by spidering Reuters.com. On the other hand, the bogus pieces were culled from a variety of publications. The fabricated news items were compiled from websites that were deemed to be untrustworthy by both the American fact-checking organization Politifact and the free encyclopedia, Wikipedia.

Multimodal detection of bogus news is another approach. In that work, writers established characteristics that may be included in the analysis. There are three types of features used in this context: statistical or semantic features extracted from text, visual features extracted from images, and features extracted from social contexts (followers, hashtags, retweets). Textual and visual indicators were employed in the given method to identify bogus news.

6. Conclusion

In this study, we introduce an exploration of fake news detection on different social media platforms. We provide the difference between unimodal and multimodal in false content information. So, we introduced most studies that implement unimodal or multimodal with classified these studies into machine learning implementation or deep learning implementation supplied with the results of each study. We compare the different models. Also, we discuss the major and important applications of fake news detection, such as fake account detection, bot detection, cyberbullying detection, and security and privacy issues in social media. Finally, we provide news insights and a challenging area of research.

Reference

[1] J. Y. Khan, M. T. I. Khondaker, S. Afroz, G. Uddin, and A. Iqbal, "A benchmark study of machine learning models for online fake news detection," *Machine Learning with Applications*, vol. 4, p. 100032, 2021.

Doi: https://doi.org/10.54216/FPA.040105
Received: February 16, 2021 Accepted: Jun 15, 2021

[2] J. Zeng, Y. Zhang, and X. Ma, "Fake news detection for epidemic emergencies via deep correlations between text and images," *Sustainable Cities and Society*, vol. 66, p. 102652, 2021.
[3] J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of massive data sets*: Cambridge university press, 2020.
[4] V. L. Rubin, N. Conroy, Y. Chen, and S. Cornwell, "Fake news or truth? using satirical cues to detect potentially misleading news," in *Proceedings of the second workshop on computational approaches to deception detection*, 2016, pp. 7-17.
[5] H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using n-gram analysis and machine learning techniques," in *International conference on intelligent, secure, and dependable systems in distributed and cloud environments*, 2017, pp. 127-138.
[6] P. Bourgonje, J. M. Schneider, and G. Rehm, "From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles," in *Proceedings of the 2017 EMNLP workshop: natural language processing meets journalism*, 2017, pp. 84-89.
[7] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," in *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, 2017, pp. 900-903.
[8] E. Fast, B. Chen, and M. S. Bernstein, "Empath: Understanding topic signals in large-scale text," in *Proceedings of the 2016 CHI conference on human factors in computing systems*, 2016, pp. 4647-4657.
[9] Y. Liu and Y.-F. B. Wu, "Fned: a deep network for fake news early detection on social media," *ACM Transactions on Information Systems (TOIS)*, vol. 38, pp. 1-33, 2020.
[10] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimedia Tools and Applications*, vol. 80, pp. 11765-11788, 2021.
[11] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, *et al.*, "Eann: Event adversarial neural networks for multimodal fake news detection," in *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*, 2018, pp. 849-857.
[12] D. Khattar, J. S. Goud, M. Gupta, and V. Varma, "Mvae: Multimodal variational autoencoder for fake news detection," in *The world wide web conference*, 2019, pp. 2915-2921.
[13] C. Song, N. Ning, Y. Zhang, and B. Wu, "A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks," *Information Processing & Management*, vol. 58, p. 102437, 2021.
[14] P. K. Roy and S. Chahar, "Fake Profile Detection on Social Networking Websites: A Comprehensive Review," *IEEE Transactions on Artificial Intelligence*, 2021.
[15] S. Khaled, N. El-Tazi, and H. M. Mokhtar, "Detecting fake accounts on social media," in *2018 IEEE international conference on big data (big data)*, 2018, pp. 3672-3681.
[16] L. Caruccio, V. Deufemia, F. Naumann, and G. Polese, "Discovering relaxed functional dependencies based on multi-attribute dominance," *IEEE Transactions on Knowledge and Data Engineering*, 2020.
[17] S. Revathi and M. Suriakala, "Profile Similarity Communication Matching Approaches for Detection of Duplicate Profiles in Online Social Network," in *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, 2018, pp. 174-182.
[18] M. Rahman, A. M. Likhon, A. Rahman, and M. H. Choudhury, "Detection of fake identities on Twitter using supervised machine learning," Brac University, 2019.
[19] N. Agarwal, S. Jabin, and S. Z. Hussain, "Analyzing real and fake users in Facebook network based on emotions," in *2019 11th International Conference on Communication Systems & Networks (COMSNETS)*, 2019, pp. 110-117.
[20] M. M. Swe and N. N. Myo, "Fake accounts detection on Twitter using a blacklist," in *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*, 2018, pp. 562-566.
[21] M. U. S. Khan, M. Ali, A. Abbas, S. U. Khan, and A. Y. Zomaya, "Segregating spammers and unsolicited bloggers from genuine experts on Twitter," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, pp. 551-560, 2016.
[22] P. V. Phad and M. Chavan, "Detecting compromised high-profile accounts on social networks," in *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2018, pp. 1-4.
[23] Z. Alom, B. Carminati, and E. Ferrari, "Detecting spam accounts on Twitter," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2018, pp. 1191-1198.

Doi: https://doi.org/10.54216/FPA.040105
Received: February 16, 2021 Accepted: Jun 15, 2021

[24] R. Aswani, A. K. Kar, and P. V. Ilavarasan, "Detection of spammers in Twitter marketing: a hybrid approach using social media analytics and bio-inspired computing," *Information Systems Frontiers*, vol. 20, pp. 515-530, 2018.

[25] K. S. Adewole, N. B. Anuar, A. Kamsin, and A. K. Sangaiah, "SMSAD: a framework for spam message and spam account detection," *Multimedia Tools and Applications*, vol. 78, pp. 3925-3960, 2019.

[26] M. Orabi, D. Mouheb, Z. Al Aghbari, and I. Kamel, "Detection of bots in social media: A systematic review," *Information Processing & Management*, vol. 57, p. 102250, 2020.

[27] L. A. Cornelissen, R. J. Barnett, P. Schoonwinkel, B. D. Eichstadt, and H. B. Magodla, "A network topology approach to bot classification," in *Proceedings of the annual conference of the South African Institute of computer scientists and information technologists*, 2018, pp. 79-88.

[28] S. Hurtado, P. Ray, and R. Marculescu, "Bot detection in Reddit political discussion," in *Proceedings of the fourth international workshop on social sensing*, 2019, pp. 30-35.

[29] N. Abu-El-Rub and A. Mueen, "Botcamp: Bot-driven interactions in social campaigns," in *The world wide web conference*, 2019, pp. 2529-2535.

[30] K. E. Daouadi, R. Z. Rebaï, and I. Amous, "Bot detection on online social networks using deep forest," in *Computer science online conference*, 2019, pp. 307-315.

[31] P. A. Chew, "Searching for unknown unknowns: unsupervised bot detection to defeat an adaptive adversary," in *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, 2018, pp. 357-366.

[32] Z. Chen and D. Subramanian, "An unsupervised approach to detect spam campaigns that use botnets on Twitter," *arXiv preprint arXiv:1804.05232*, 2018.

[33] P. Shi, Z. Zhang, and K.-K. R. Choo, "Detecting malicious social bots based on clickstream sequences," *IEEE Access*, vol. 7, pp. 28855-28862, 2019.

[34] A. Dorri, M. Abadi, and M. Dadfarnia, "SocialBotHunter: Botnet detection in Twitter-like social networking services using semi-supervised collective classification," in *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, 2018, pp. 496-503.

[35] A. Alarifi, M. Alsaleh, and A. Al-Salman, "Twitter Turing test: Identifying social machines," *Information Sciences*, vol. 372, pp. 332-346, 2016.

[36] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race," in *Proceedings of the 26th international conference on world wide web companion*, 2017, pp. 963-972.

[37] J. Echeverria and S. Zhou, "Discovery, retrieval, and analysis of the star wars' botnet in Twitter," in *Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining 2017*, 2017, pp. 1-8.

[38] K. Lee, B. D. Eoff, and J. Caverlee, "Seven months with the devils: A long-term study of content polluters on Twitter," in *Fifth international AAAI conference on weblogs and social media*, 2011.

[39] F. Morstatter, L. Wu, T. H. Nazer, K. M. Carley, and H. Liu, "A new approach to bot detection: striking the balance between precision and recall," in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2016, pp. 533-540.

Doi: https://doi.org/10.54216/FPA.040105
Received: February 16, 2021 Accepted: Jun 15, 2021