International Journal of Information Security (2023) 22:177–212
https://doi.org/10.1007/s10207-022-00625-3

SURVEY

A review on fake news detection 3T's: typology, time of detection, taxonomies

Shubhangi Rastogi[1] · Divya Bansal[1]

Published online: 15 November 2022

Abstract

Fake news has become an industry on its own, where users paid to write fake news and create clickbait content to allure the audience. Apparently, the detection of fake news is a crucial problem and several studies have proposed machine-learning-based techniques to combat fake news. Existing surveys present the review of proposed solutions, while this survey presents several aspects that are required to be considered before designing an effective solution. To this aim, we provide a comprehensive overview of false news detection. The survey presents (1) a clarity to problem definition by explaining different types of false information (like fake news, rumor, clickbait, satire, and hoax) with real-life examples, (2) a list of actors involved in spreading false information, (3) actions taken by service providers, (4) a list of publicly available datasets for fake news in three different formats, i.e., texts, images, and videos, (5) a novel three-phase detection model based on the time of detection, (6) four different taxonomies to classify research based on new-fangled viewpoints in order to provide a succinct roadmap for future, and (7) key bibliometric indicators. In a nutshell, the survey focuses on three key aspects represented as the three T's: Typology of false information, Time of detection, and Taxonomies to classify research. Finally, by reviewing and summarizing several studies on fake news, we outline some potential research directions.

Keywords Fake news · Typology · Methodology · Survey · Satire · Datasets

1 Introduction

Today, the web, social media, and other forums have become the primary source of information over traditional media [9]. The freedom of expression, spontaneous and real-time information provided by social media platforms make it a popular topic of interest, especially among the younger generation. Consumers use these platforms worldwide to access news related to everything from celebrities to politics and often take for granted whether the news is authentic or not [128]. Essentially, the motive of social media platforms is to get users engaged to earn business revenues rather than providing factual information. The problem of false information was best emphasized during the US presidential election in 2016, which remains under investigation. Also, the covid-19 pandemic has shown various instances of fake news [163] like political propaganda, health-threatening false news, hate speech, etc. Fake information is spreading faster than the virus itself, which led to the introduction of a new term, '*infodemic*.' Identification of fake news is a complex problem because it is swiftly becoming an industry on its own, where users paid to write conspiracy theories and create clickbait content to allure the audience. Thus, fake news detection has become an emerging research area and several studies have been done to provide solutions for identifying fake information.

Due to the explosion of information on social media, the manual fact-checking of each post is impossible. Manual fact-checking by humans is time-consuming and subject to human bias. The alternative approach is to leverage machine learning algorithms in order to automate the process of fake news detection. However, machine learning-based solutions impose a few limitations, such as obtaining a large training dataset and selecting suitable features which can best capture the deception. The strong literature exists to present different proposed solutions based on machine-learning algorithms for the detection of false information [2]. Nonetheless, this

✉ Shubhangi Rastogi
shubhangi05.rastogi@gmail.com

Divya Bansal
divya@pec.edu.in

1 Punjab Engineering College (Deemed to be University), Chandigarh, India

survey provides different aspects required to be considered before designing an effective solution. This survey highlights this research gap to understand these aspects of the problem before designing the methodology. Therefore, the goal of this study is multipronged and covers important elements required to be considered before designing the methodology for fake news detection. The paper also provides a bibliometric analysis to find potential areas of research in fake news. The elements are not sole accountable for the identification of false information, but rather one of many indicators.

Thus, the survey has been done with two broad objectives. The first objective is to provide a comprehensive overview of current studies in this area by highlighting multiple directions valuable before designing the machine-learning-based methodology. Based on our explication, we introduced a typology of false information, a three-phase model based on time of detection, four taxonomies on different indicators, existing datasets, actors involved in spreading false information, and actions taken by service providers. Supervised machine learning approaches are dependent on externally supplied data [71], and therefore, these identified elements (or directions) play an important role in improving the performance of such algorithms. The second objective presents a bibliometric analysis that will help the readers to find out the top organizations, funding agencies, journals, keywords in articles, etc. in this research area. The demographic spread may also assist in interdisciplinary research. Since the topic of fake news, detection is timely and has a lot of scope in research; thus, this article will help new researchers to develop an interest by understanding various perspectives and follow a road path for their future work. To the best of our knowledge, no existing survey covers these directions which are important especially for new researchers to obtain a comprehensive overview of the domain.

To summarize, we shed light on the following research questions: (1) What are the different types of false information? The paper presents a clear problem characterization with apt definitions and examples of each concept related to fake news. (2) Who are the actors that spread false information? (3) What actions have been taken by the service providers to mitigate fake news? (4) What are the available datasets for text, image, and video data formats? (5) How time is an important factor in the detection of fake news? (6) What are the novel viewpoints of classifying research in this domain in order to provide a useful future roadmap? (7) What are the foremost journals and organizations in this research area? The literature shows several survey articles as explained in Sect. 1.1; however, this survey aims to outline comprehensive research on fake news detection with a number of contributions highlighted in Sect. 1.2.

1.1 Survey archetypes

The literature contains a variety of survey papers on the detection of fake news and related terms. We have segregated different recent survey papers into four categories depending on the type of research design. This section aims to showcase the existing surveys' style to differentiate our proposed survey from the existing surveys in this research domain. According to the method of reviewing existing studies, four types of surveys are:

(a) **Type I**: Misinformation, Disinformation, and Mal-information
Some surveys focused on three main categories of information disorder: misinformation, disinformation, and mal-information. Misinformation and disinformation have been used interchangeably in much of the discourse on fake news. However, the two categories differ in terms of the degree of falseness and intent to harm. Misinformation is unverified news, but the source/spreader is unaware, and the intention is not to harm the public, while disinformation is unauthentic news to mislead the audience and the source/spreader knows it is false. The third category, mal-information is the deliberate dissemination of news (which is real) in order to harm a person, specific organization, or country, e.g., leaking private information, or disclosing one's sexual orientation without public interest justification. Therefore, mal-information is not fake information but unethical. Figure 1 has been popularly used by the literature.

(b) **Type II**: False information typology: Fake news, satire, rumor, clickbait, hoax
In the literature, there exist survey papers highlighting unclarity in the problem definition. There are different definitions of fake news given by researchers and psychologists. Fake news is the most popular information disorder, but it is different from other types. For example, Satire news is for entertainment and hides some humor inside. Studies consider fake news and satirical news in the same boat; however, the whole purpose of both types is completely different. In a similar way, rumors, clickbait, and hoaxes have different agendas and influences on the audience. Section 2 contains the detailed eyeshot of this problem.

(c) **Type III**: Research Approach for fake information detection:
Majorly, surveys generally classify research based on the methodology. The generic methodological framework is shown in Fig. 2, which shows a roadmap of existing solutions in the literature. In particular, studies have focused on three foremost steps: (1) data collection, (2) feature extraction, and (3) Classification technique.

The novelty in research works exists in these steps. Data collection is a challenging task, and researchers have presented different ways to extract data as well as publicly available datasets. Due to the absence of a benchmark dataset for fake news detection, researchers are required to apply different classification models to find the best one suitable for the given dataset. The generic methodology presents different step-by-step paths that have been proposed in the literature. The classification techniques are coarsely grouped into two categories, machine learning [103] and deep learning-based methods [104] [102] [98]. Machine learning has shown the promising results in this application area of classification. The traditional machine learning approaches are based on handcrafted feature extraction. The generic methodology has shown some handcrafted features, such as linguistic features and propagation patterns, which are prominent in the literature. However, diagnosing relevant features to best capture the deception imposes another challenge. It is time-consuming and may result in biased features, especially in such domains, as fake news detection. Therefore, deep learning approaches have gained popularity in solving such critical problems. Deep learning models can learn hidden representations using neural networks, so the extraction of handcrafted features is not required here. Thus, the focus is transferred from modeling relevant features to modeling a network itself. This methodology also highlights different machine learning and deep learning algorithms with their categorization. The methodology has been designed to give readers an overview of the existing steps followed by the researchers in order to understand and implement further solutions. Each path presents a research direction or approach to solve the problem of fake news classification. Furthermore, several studies have employed additional approaches such as the Hawkes process and anomaly detection. In a nutshell, several surveys have been proposed based on the research approaches followed by the researchers to identify fake news.

(d) **Type IV**: Perspectives
Several surveys reviewed the research based on perspectives. There are four perspectives given the literature for automatic detection of fake news: the unfactual *knowledge* it conveys, its *style of writing* or content-based, its *propagation* patterns or social, and the *credibility* of its source. Figure 3 shows the four perspectives with the features used in each. Several research works have been studied in order to describe these perspectives and corresponding features diagrammatically [196] [149].

Table 1 highlights several archetypes of survey papers. The existing survey papers present extensive insights into the research area. However, we present this survey paper to highlight some innovative aspects: (1) The survey provides simple and clear definitions of fake news and related concepts with the help of real-life examples. We have deeply analyzed various real-life instances to define fake and related concepts. Also, it explains approaches used in the state of the art for each type. This will help the readers to define boundaries between different types of false information. (2) This survey groups actors involved in spreading misinformation. Several instances of fake news spread on digital sources have been analyzed to understand the role of the user posting it in order to group them into different categories. (3) Most people use social media platforms to access news; thus, these platforms also take appropriate actions to mitigate fake news. Our survey presents some recent actions taken by social media platforms. (4) Related survey papers review existing datasets but our survey reviews datasets in this domain based on the content type, i.e., text, image, and video. (5) Our survey segregates recent surveys into four types; thus, it will help readers compare and analyze related surveys. (6) Our survey presents taxonomies based on domain, data type, and platform. (7) Our survey presents a novel approach to segregating research based on detection time. (8) To end, this survey provides a bibliometric analysis. To the best of our knowledge, no survey includes these aspects.

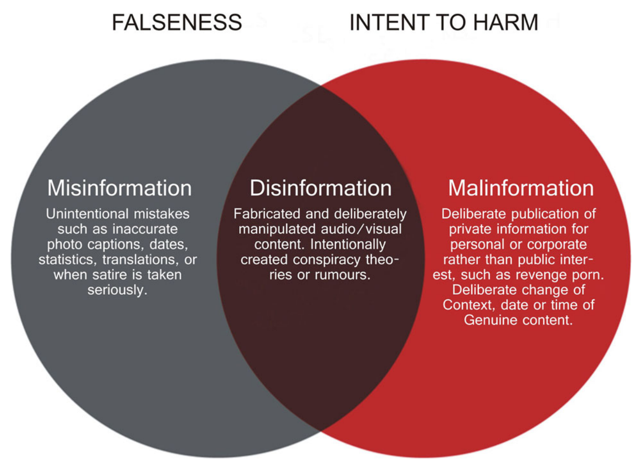


Fig. 1 Venn diagram for misinformation, disinformation, and malinformation Source: Wardle et al. [179]

1.2 Organization of the paper and key contributions

We have presented the typology in information disorder in Sect. 2 by explaining scholarly studies on fake news and related false information types. The major source of false information is social media platforms; thus, we have provided statistics of popular social media platforms which had fake news history with the actions taken by the service providers to control fake news (Sect. 3.2). Section 4 provides the details of publicly available datasets considering three data formats

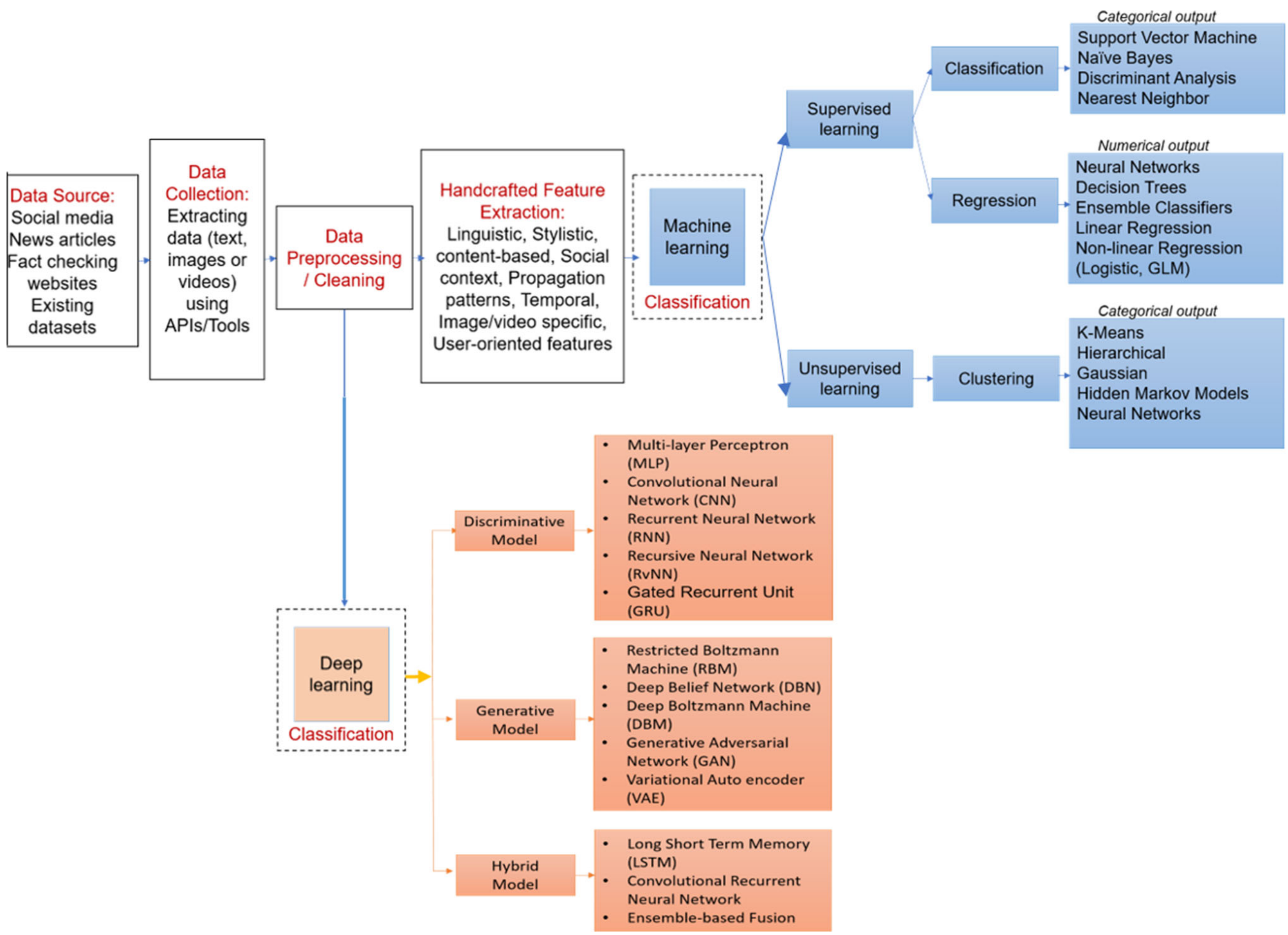


Fig. 2 Generic methodology for detection of fake news

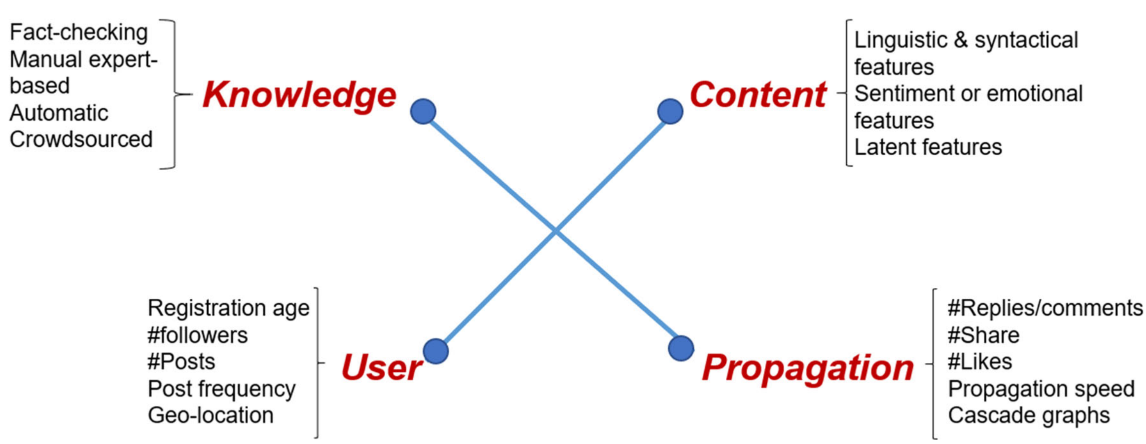


Fig. 3 Four Perspectives of Fake news detection

such as text, images, and videos. Section 5 gives a novel 'three-phase model' for the detection of fake news depending on the time of detection. The paper presented several taxonomies based on domain, features per misinformation type, misinformation data type, and platform (Sect. 6). Furthermore, Sect. 7 delivers the statistical analysis of several studies and also presents the journal ranking and visualization of bibliometric analysis based on *'InCites'*(website for

Table 1 Detailed comparison of existing surveys

Ref.	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
[179]	✓							
[12]	✓				I			
[100]	✓		✓	✓	II,III			
[187]	✓	✓			II			
[188]				✓	IV			
[196]	✓				IV			
[134]	✓				II			
[20]	✓			✓	II,III,			
[78]	✓	✓	✓		II,III			
[3]	✓				IV			
[149]	✓			✓	IV			
[167]			✓	✓	III	✓		
[170]	✓			✓	III			
Proposed Survey	✓	✓	✓	✓	–	✓	✓	✓

Q1: Define Fake News and related concepts? Q2: If and how they group actors of fake news? Q3: If they investigate social media platforms? Q4: If they review existing datasets? Q5: What is the type of survey? Q6: If it covers news domains, multimodal research, multiplatform, etc. Q7: If it covers timeline of news for detection Q8: If it provides bibliometric analysis

ranking Science citation index (SCI) articles). In Sect. 8, we discuss open issues in fake news detection to facilitate efficient research in this domain. Finally, we conclude in Sect. 9. The key contributions of the work are as under:

1. Puts forward the veracity and variety of false information by providing a clear definition of different types such as fake news, hoax, rumor, clickbait, and satire. A real-life example has been provided for each type to provide clarity in problem definition. Also, our survey classifies various studies in the literature focusing on a particular type.
2. Although recent studies highlighted various approaches proposed by researchers to detect fake news, we provide a statistic of measures taken by service providers to combat and mitigate fake news.
3. Publicly available datasets are outlined considering data formats, i.e., text, images, and videos.
4. Current surveys have mostly reviewed research with four designs described in Sect. 1.1. Our survey presents a novel way to look over the problem, i.e., the time of detection. The existing studies are based on the assumption that they have all the lifecycle data. But it depends on the time when the news was spread. The survey gives a 'three-phase model' for early, mid-, and late detection of false information.
5. Our survey also provides four different taxonomies in order to offer a succinct roadmap for future work. First, the taxonomy classifies research into five different domains along with the articles based on a cross-domain approach. Second, the taxonomy presents a 2D view to identifying prominent features in the literature as per the type of false information. Third, the taxonomy classifies studies based on data format under consideration as well as research on multimodal data. Fourth, the taxonomy classifies research based on the platform such as social media, Wikipedia, and fact-checking websites, including multi-platform. To the best of our knowledge, no survey has classified the research in these ways.
6. The study also presents some key bibliometric indicators like highly cited papers, publication trends of more than a decade, journal citations, etc.
7. The state of the art and research gap specified in the paper help in deciding the future path to combat this rampant problem of fake news.

2 Typology of information disorder

This section explains the veracity and variety of misinformation and gives a proper definition with real-life examples to differentiate the types as summarized in Table 2. There is an overlay in the types of false information, i.e., one type of false news may fall into multiple sorts. For instance, a rumor may use clickbait approaches to allure the readers and increase its reach. The reason behind this overlay is that the creators are also evolving their style or pattern of writing false information; thus, obtaining clear boundaries between these types is complex. But we need to comprehend the term fake news and differentiate it from other concepts, like satire, rumor, clickbait, hoaxes, etc. Several research studies have presented definitions of the concepts associated with 'fake news.' Our survey simplifies the problem of differentiating these terms by providing a typology with real-life examples for each type. Figure 4 outlines different illustrations for each false information type. To this end, Table 3 illustrates the overall categorizations of the state of the art in order to learn the variances in detecting different types of false information in terms of input data type, research technique, dataset technique, results, and limitations of the research. This comparison of fake news and related concepts based on the existing work in the literature shows the use of machine learning approaches to automate the detection of false information. This, in turn, is useful for the readers to design an effective machine learning-based methodology based on the given type of false information. For example, satirical news is less harmful than fake news, thus requires a different approach for detection. To further support this, we have also presented the prominent features in the literature

Table 2 Misinformation Typology

Type	Definition	Example
Fake news	Completely false stories	Hillary Clinton adopted an alien baby
Propaganda	Special instance of fabricated stories that aim to harm a particular party	BlackLivesMatter, Syria airstrikes 2018
Conspiracy theories	Stories that try to explain an event by invoking conspiracy without proof	Pizzagate theory, Seth Rich
Hoaxes	Half-truth or humor	False death of celebrities, April fools' day events
Biased	An alt-right echo chamber	4chan's/pol/board
Rumors	Whose truthfulness is ambiguous or never confirmed	Chennai Floods, 2013 Boston Marathon Bombings
Clickbait	Misleading headlines and is least severe types	Yellow journalism
Satire news	Irony + humor	Sites that post satire news: TheOnion and SatireWire

Fig. 4 Illustrations of different types of information disorders **a** Fake news spread on Twitter during 2016 US presidential election **b** One of the hoaxes related to Sonali Bendre's death spread over social media (Indian Politician Mr. Ram Kadam posted this tweet in Marathi and then deleted it after backlash on Twitter) **c** Rumors spread during Chennai Floods **d** Popular Clickbait on social media **e** Satire news related to IAF Balakot airstrike

for each type of information required by machine-learning classifiers in Sect. 6.2.

2.1 Fake news

Fake news is fabricated news that is completely unreal. There is no truth behind the fake news, and it is verifiably false [196]. This kind of information disorder is indigenous and exists from the time of world wars and earlier. It is difficult to create fake news which cannot be deciphered easily. Therefore, creators reform the existing fake news, which makes it look like genuine news. Figueira et al. [45] highlighted the current state of fake news and created a fact-checking algorithm including three W's: 'who, where, when.' Recently, Vishwakarma et al. [171] proposed a model to analyze the veracity of information present on social media platforms in the form of an image. Essentially, the study used text extraction from image technique to classify news with an image as fake or real. The paper has given a good set of examples of fake and real images. Perez et al. [120] presented a twofold approach in order to detect fake news. First, two novel datasets were constructed based on several domains and, second, conducted a set of learning experiments to build an automatic fake news detector with an accuracy of 76%. Primarily, researchers have been using different features to detect fake news. Furthermore, Reis et al. [130] explored the existing features in the literature and also presented some novel features to accurately detect fake news. The five features proposed are bias, credibility, domain location, engagement, and temporal patterns. The features are not only generated from the content present in the news rather some features can also be mined too. Olivieri et al. [116] proposed a methodology to create task-generic features using metadata obtained from Google custom-search API. The features created are statement domain scores and similarities for titles and snippets by collecting metadata corresponding to the top 20 results of Google search. Similarly, Ahmed et al. [2] investigated two different feature extraction techniques, namely term frequency (TF) and term frequency–inverted document frequency (TF-IDF) over n-gram analysis and machine learning models. They found 92% accuracy with linear support vector machine (LSVM) using TF-IDF.

2.2 Hoax

Hoaxes are half-truths, different from fake news, which is a full-blown lie. The fake news menace is more than a hoax because fake news affects the public and is like an epidemic disease, while hoaxes are made for fun and get exhausted after one step. Examples of hoaxes generally include the sudden death of a celebrity. The literature on hoaxes is not as wide as fake news, and the reason may be the aftereffects are not comparably less serious. Researchers of Lancaster University, UK, examined the practical jokes shared on 1 April, which is known as April fools' day, and observed that this data could be used as a hoax dataset [33]. Also, to construct a set of features mainly, linguistic features from the past research to detect deception, humor, and fake news. They observed that the hoax dataset could be used to detect fake news based on a similar feature set. Fauzi et al. [43] highlighted a few features to detect news that has a tendency toward hoaxes. Sentiment analysis and Tweets containing provocation, feud, and anxiety words are identified, followed by the 'SVM' machine learning model to detect hoax possibilities. In another work, Situngkir et al. [158] reported the propagation of hoaxes on the social media platform Twitter. In this paper, the case study of the death of a public figure in Indonesia has been used to examine the epidemy of hoaxes over social media. Tacchini et al. [160] presented a list of Facebook pages divided into two categories: scientific news sources and conspiracy news sources to collect datasets for non-hoaxes and hoaxes posts, respectively. Furthermore, several classification experiments have been done to find different useful observations. Formerly, Vukovic et al. [174] stated that generally, hoaxes are harmless, but they may harm someone's image by deceiving the readers. The dataset is constructed using real email messages and real email hoaxes. A hoax detection system proposed by the researchers is successful to some extent. Also, the same system can be used to detect SMS hoaxes. Essentially, literature considers different sources to collect hoax datasets. Similarly, Kumar et al. [79] highlighted the role of Wikipedia to spread hoaxes. Wikipedia, being an open crowdsourced platform, has the power to attract false information propagation. The study used already flagged articles by Wikipedia editors as a Hoax dataset for future predictions by exploiting the similar feature pattern shared by hoaxes.

2.3 Rumor

Rumors are ambiguous stories whose truthfulness never gets confirmed or it gets confirmed after a long period of time till that damage gets already been done. The strong literature exists in this domain primarily on the basis of different platforms used to spread rumors. Rumors exist in two forms breaking news rumors and long-lasting rumors. Alkhodair et al. [7] worked on detecting real-time Twitter stream of breaking news rumors related to emerging topics. The authors used the publicly accessible dataset 'PHEME' and trained the model using deep learning and machine learning. Recurrent neural network–long short-term memory (RNN-LSTM) has been used along with word embeddings. Also, compare this model with non-sequential classifiers by considering two feature sets, namely, content-based and social-based features. Moreover, the authors demonstrated the performance of their model on a real-time Twitter stream of breaking news. Rumor detection in the English language is a flowing research area,

Table 3 Literature review

Refs.	Purpose	Input data form	False information type	Datasets and Size	Technique	Result	Limitation/future scope
[171]	Validate the veracity of information using credibility of top 15 google search	Image +Text	Fake news	Google Images, the Onion, and Kaggle	Credibility of the top 15 Google search using page rank algorithm	85% accuracy	Improvement in the process of entity extraction for images
[120]	Develop two novel datasets & Use linguistic features to classify fake and real	Text	Fake news	FakeNewsAMT, Celebrity 480 & 500	Machine learning classifier: SVM	76% accuracy	Required to apply hybrid decision models with fact verification and data-driven machine learning judgments
[59]	Fake and satire news are distinguishable using stylistic features of the title.	Text	Fake + Satire	Buzzfeed election data + real fake satire websites + Burfoot and Baldwin data 120,224,4233	Machine learning SVM	91% accuracy	Limited dataset size
[190]	Propose a clickbait convolutional neural network (CBCNN)	Text	Clickbait	Scraped Chinese news headlines 14,922	Word2vec + CNN	80.5% accuracy	Maximum length of the headline is limited causing information loss for long headlines

Table 3 continued

Refs.	Purpose	Input data form	False information type	Datasets and Size	Technique	Result	Limitation/future scope
[168]	Detection of Clickbait videos on YouTube using cognitive evidences	Video	Clickbait	Crawled YouTube videos 987	Set of supervised models	J48 98.89%	Dataset limitation
[96]	Novel rumor detection approach using entity recognition, sentence reconfiguration and ordinary differential equation network	Text	Rumor	Publicly available dataset by Ma et al. [95] 1,101,985	GRU, LSTM, CNN, ODE	85.89% F1 score	Heavy time cost
[86]	Multimodal approach for satire detection using textual and visual features	Text + Image	Satire	Scraped data from satirical and regular websites 10,000	Language BERT (ViLBERT)	93.80% accuracy	Lack of image forensics methods
[185]	Hoax news detection using reader feedback under two conditions with and without URL	Text	Hoax	Crawled news data from a hoax website 250	Machine learning classifiers SVM, C4.5, Naive Bayes	0.95 F1 measure	Tested on limited classifiers and include sentiment analysis in future

Table 3 continued

Refs.	Purpose	Input data form	False information type	Datasets and Size	Technique	Result	Limitation/future scope
[117]	Develop fake news detection model named Cross-SEAN during Covid-19	Text	Fake news	Scraped Covid-19 related tweets 46.26k	Semi-supervised neural attention model	95.4% accuracy	May exist bias in the external knowledge and absence of early detection
[4]	Comparison of different feature sets in order to detect sarcastic articles	Text	Sarcasm	Scraped data using twitter API	12 Machine learning algorithms: gradient boosting, Gaussian Naïve Bayes, AdaBoost	80% accuracy highest with gradient boosting	Limited number of features
[2]	Fake news detection using n-grams analysis	Text	Fake news	News scraped from Reuters and Kaggle	SVM machine learning	92% accuracy	Required to test on real-time scraped data
[160]	Classify Facebook posts into hoaxes and non-hoaxes based on users who 'liked' them	Text	Hoax	Scraped Facebook posts 15500	Logistic regression & Boolean label crowdsourcing	99% accuracy	Dataset limitation
[10]	Rumor detection in Arabic tweets using content and user-based features	Text	Rumor	Scraped using Twitter API 271000	Semi-supervised & unsupervised model using expectation-maximization (E-M)	78.6% accuracy	Maximize unlabeled data to improve accuracy

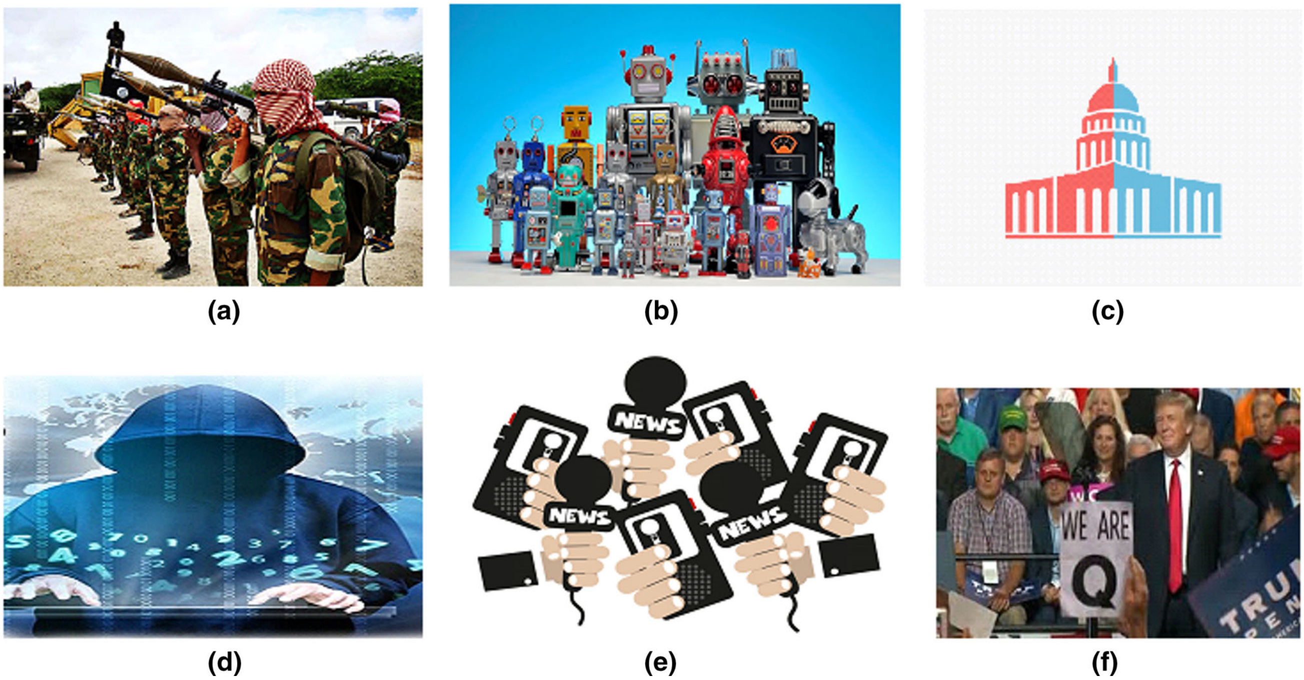


Fig. 5 Actors involved in spreading misinformation: **a** Terrorist organizations, **b** Bots, **c** Government, **d** State-sponsored trolls, **e** journalists, **f** Conspiracy theorists 'true believers'

but there is a need to explore other languages in this field. Alzanin et al. [10] extracted 271000 rumor and non-rumor tweets using Twitter API, and rumor topics were obtained from anti-rumors authority. After preprocessing, content and user-based features have been extracted, and finally, semi-supervised expectation-maximization (E-M) compared with supervised Gaussian NB. Rumors are popular in the health-related domain also. Sicilia et al. [151] presented a novel approach including new features, namely influence potential and network characteristics. Furthermore, different feature selection methods were explored using a few classification methods for rumor detection and finally validated the system as well as these features on a real Twitter dataset by achieving 90% of accuracy. Rumor detection systems can be used to compare different social media platforms. Priya et al. [123] compared Reddit and Twitter using different features based on content and social influence. It was found that Reddit is better for a conceptual overview, while Twitter is better for evolutionary analysis because of the size constraint and longer span of time of Twitter microblogs.

2.4 Clickbait

Clickbait is misleading headlines that make the audience crave the story. Essentially, these are attractive headlines, but the story behind these headlines is completely different. On social media platforms like Facebook, Twitter, etc. Daoud et al. [31] proposed an effective approach for clickbait detection based on supervised machine learning techniques. Some important features for detecting clickbait are similarity between text and title, the formality of the language used, readability, and the bag of words. Since the number of features is large, recursive feature elimination was used on the SVM classifier to obtain an accuracy of 79%. The literature in clickbait detection focuses on exploring new features. According to Biyani et al. [18], common features for fake news or spam detection like link structure, blacklists of URLs hosts, and IPs are not advantageous for clickbait detection. Clickbait detection requires feature engineering based on content, similarity, and informality kind of parameters. Clickbait detection is gaining attention in research, but the methods majorly require aggressive feature engineering. Zheng et al. [190] found that a convolutional neural network performs better than traditional machine learning models considering the word-sequence information and learning word meanings from the whole dataset. Apart from microblogging platforms like Twitter, online video sharing platforms like YouTube are also getting popular. Clickbait video, whose content is not related to the title, is an emerging research area. Shang et al. [147] proposed a novel content-free approach named Online Video Clickbait Protector (OVCP) to effectively detect clickbait videos by analyzing the comments shared on the video by the audience.

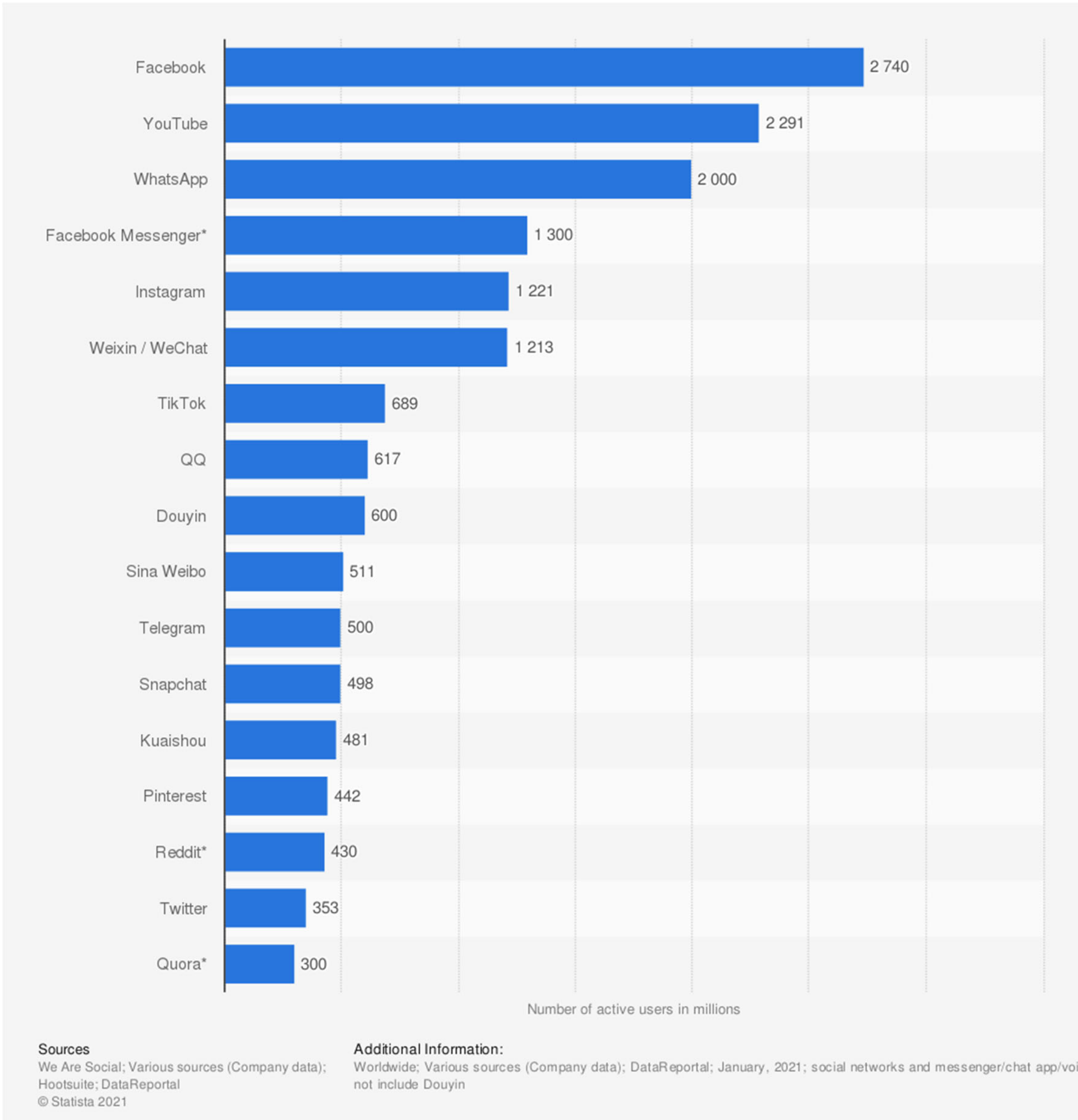


Fig. 6 Most popular social networks worldwide as of January 2021, ranked by number of active users (in millions)

2.5 Satire

Satirical news is created and published for entertainment purposes. In general, there should be a separate room for satirical news and should not be mixed with fake news. Satirical news is different from other false information types in the sense that it intentionally signals its deceptiveness, while other types try to develop a false sense of truth in the mind of the audience [127]. Therefore, any average knowledgeable person can easily distinguish between satire and fake news. The problem persists when readers start taking satire news as legitimate news. The literature emphasize how the approaches to detect satire are different from other deceptive news. Rubin et al. [135] developed a feature set (absurdity, grammar, and punctuation) that can best capture the deception in satire news. The dataset was obtained from satirical sites and legitimate sites, and multiple experiments were performed to identify this best performing feature set with an accuracy of 82%. The important observations were satire news is longer than legitimate, satire news is commonly grammatical incorrect, and use more punctuations. Data collection is a problem in any detection of any misinformation, but this work only considered the direct satire sources and legitimate sources to build the corpus. However, satire news is more popular on Twitter nowadays, which was missing in their work. Ahuja et al. [4] proposed a simple approach for data collection from Twitter. Tweets with direct positive and negative sentiments were crawled using keywords #cheerful, #happy and

#sorrow, #angry, respectively. While collecting satire data, keywords like #sarcasm #sarcastic were used. Moreover, the authors checked twelve classifiers on different feature sets obtained and compared them in terms of accuracy. Thu et al. [161] used a lexicon-based approach to extract different emotional features. Three lexicons, namely EmoLex, VADER, and SentiNet, were exploited to extract emotional features, and finally, data are classified into two categories, satire, and non-satire using random forest and SVM distributed over different emotions like anger, trust, and many more. Horne et al. [59] compared real, fake, and satire news with respect to stylistic, complexity, and psychological features on three standard datasets. Authors found that fake news, in most cases, is more similar to satire than real news. Also, titles of fake news use notably fewer stop-words and nouns while using more proper nouns and verb phrases.

3 Actors and actions

3.1 Actors involved in spreading false information

False information is hard to fight, in part because it circulates for all kinds of reasons. Sometimes its bad actors churning out fake news in a quest for internet clicks and advertising revenues, there are 'troll farms' that create misleading stories, and other times, the recipients of false information are driving its propagation. Thus, it is important to highlight the actors that are involved in the circulation of false information. We have identified a few actors in Fig. 5. The popular actors involved in false information are terrorist organizations (e.g., ISIS), bots (autonomous software to repost false data) [40], governments (historical instances of the Russian government in US elections), trolls (posts provoking content), journalists (modify/exaggerate a narrative to make it attractive), and true believers. (They strongly believe the false story is true and spread it.) Similar to false information typology, the intersection may also occur in actors.

3.2 Actions taken by service providers

People use social media platforms not only to get connected with friends and family but also to access news. Users on social media platforms post an enormous amount of data every day. Globally, over 3.6 billion people use social media, the statistics of users per platform is shown in Fig. 6 [1]. These platforms provide freedom of expression to the users in a democracy. However, the main motive of these platforms is to get users engaged to earn business revenues rather than providing them with factual information. Also, recommendation system algorithms are running behind social media platforms due to which users see their point of interest without dwelling into facts. Since the majority of youngsters nowadays follow these social media platforms to access news, they get trapped in propaganda rather than following authentic news. Government authorities of different nations have been asking these platforms to take necessary actions to control the dissemination of fake news. For instance, Twitter has recently suspended accounts of Donald Trump (former US president) and Kangana Ranaut (Indian celebrity) due to their hate and provoking posts. Facebook also flags false information. Crawford et al. have described the working of flag (an annotation to offensive or problematic content) in different social media platforms [28]. However, state of the art shows that data are extracted using digital source API and then annotated manually by domain experts for training on machine learning models. Table 4 [2] [3] [4] provides the statistics of popular social media platforms which had fake news history along with their primary features, measures taken by these platforms to stop fake news and how the corresponding annotations can be used by researchers.

This information is useful to the researchers to modify their methodology with the amendments done by the service providers.

4 Existing datasets and tools

Dataset creation is a challenging task, and researchers have explored various online information sources to extract useful data given in Table 5. Due to privacy restrictions on extracting data from online information sites, obtaining a dataset for academic research is not a straightforward task. One way to overcome this issue is to purchase data from these platforms or other crowdsourcing websites. Another way is to use existing datasets in the literature that satisfies one's research requirements. Table 7 lists the details of various datasets widely used in the literature. The comparison of existing datasets has been done based on language, label (binary, multi-class, or numeric), class distribution (balanced or imbalanced), and annotation method (expert-based or crowdsourcing-based). Such information can significantly help researchers select proper datasets for their research, which can be multilingual or focus on low-resource languages. It can also help to select evaluation metrics and evaluate the annotation quality. We have analyzed various papers which have used these datasets and included the highest accuracy reported so far for the respective datasets in the

[1] https://www.statista.com/

[2] https://www.statista.com/

[3] https://www.adobe.com/express/learn/blog/top-social-media-sites

[4] Times: https://timesofindia.indiatimes.com/ Tribune: https://www.tribuneindia.com

Table 4 Social media Platforms: facts and actions taken to control fake content

OSN	Founded year	No. of active users	Primary features	Actions to stop fake news	Annotated data for research
Twitter	2006	321 million	Registered users can post, like and retweet tweets, but unregistered users can only read them; multilingual; Label falsely claim;	Twitter forbids posts that manipulate elections; adding warnings and restrictions	Annotations have been added to the tweet object from all v2 endpoints in two forms: Entity & Context annotation[a]. Recently, Twitter launched a streaming endpoint to access COVID-19 annotated data.
Facebook	2004	2.70 billion	Registration required to do any activity	Disrupting economic incentives; building new products; Flagging fake news; Banned in few countries	Researchers can access public Facebook data flagged by its fact-checkers [47].
WhatsApp	2009	1.6 billion	Freeware, cross-platform messaging; voice and video calls, and share images, documents, user locations, and other media	WhatsApp has launched a nationwide campaign called ‘Share Joy, Not Rumors’; partnering with fact-checking websites, IFCNs helpline numbers and digital literacy NGOs	Researchers can access WhatsApp public data labeled as misinformation by professional fact-checking agencies [129].
Skype	2003	300 million	Providing video chat and voice calls; instant messaging services; video conference calls.	Provision to report computer security vulnerability	No convincing evidence available in the literature to leverage annotated data by Skype for false information detection.
YouTube	2005	1.9 billion	Online video-sharing platform; allows users to upload, view, rate, share, add to playlists, report, comment on videos, and subscribe to other users	Prioritize ‘authoritative voices’; launched top news shelf on YouTube homepage; Improved ranking systems and machine learning classifier	Researchers can utilize videos reported by YouTube and a record for data verification is available even if it is removed [62] [139].

[a] https://developer.twitter.com/en/docs/twitter-api/annotations/overview

Table 4 continued

OSN	Founded year	No. of active users	Primary features	Actions to stop fake news	Annotated data for research
Reddit	2005	330 million	Social news aggregation, web content rating, and discussion website; multilingual	Apply a quarantine to communities that contain hoax or misinformation content, removing them from search results like subreddit r/NoNewNormal, warning users and requiring explicit options to see the content.	Scraping data without Reddit's prior consent is prohibited for few subreddits like r/depression and r/SuicideWatch [124].
Instagram	2010	1 billion	Photo and video-sharing social networking service owned by Facebook;	Working with third-party fact-checkers to help identify, review, and label false information.	Facebook published heavily annotated research on Instagram's toxicity after getting reported by WSJ[a].
Snapchat	2011	287 million	Photo sharing, Instant messaging, Video chat, Multimedia pictures, and messages are usually only available for a short time;	Hire journalists and fact check everything	Snapchat is almost free of fake news [24]; thus, annotated data is not available for researchers.
Tumblr	2007	642 million	Allows users to post multimedia and other content to a short-form blog; Users can follow other users' blogs	Launched its internet literacy campaign World Wide What.	Researchers can utilize flagged Tumblr posts; for example, Wired researchers analyzed posts erroneously removed by Tumblr [155].
LinkedIn	2003	645 million 706(2020)	Professional networking, including employers posting jobs and job seekers, posting their CVs	Gives users option to flag inappropriate or fake profiles on its platform; It more relies on users than AI	LinkedIn flags the user trying to extract user profile data as bot. Thus, limited studies available in the literature to utilize its flagged data for misinformation detection.

[a] https://about.fb.com/news/2021/09/research-teen-well-being-and-instagram/

Table 5 Online information sources for data collection

Social media platforms	Twitter, Facebook, Instagram, Reddit, 4chan, 8chan, Sina Weibo, WhatsApp, YouTube
Popular news sources	Wall Street Journal, The Economist, BBC, NPR, ABC, CBS, India Today, The Guardian, News18, Times of India
Fake sources	Ending the fed, True Pundit, abcnews.com.co, DC Gazette, Liberty writersnews, Before its News, Infowars, Real News Right Now
Satire sources	Faking News, The UnReal Times, Farzinews, Newsthatmattersnot, The Onion, Huff Post Satire, Borowitz Report, The Beaverton, Satire Wire
Fact-checking sources	Snopes, Politifact, Altnews, Boom, SMHoaxSlayer, Factly, Facthunt

table. Moreover, there are various fact-checkers tools available online to determine the credibility of digital news. Table 6 outlines few popular tools for fact-checking covering different grounds such as NLP-based, bot detection, gamified version, and blockchain-based.

5 The time of detection is important

The four perspectives for the detection of fake news, such as knowledge, style, propagation, and source-based, are independent. The literature shows a rich set of features under each perspective that works best for the designed environment. However, as per the no free lunch theorem, a general methodology is impossible for every scenario. Therefore, recent advancements show various studies based on different perspectives to detect fake news. Whereas, this paper gives a novel way to look over the problem, i.e., first analyze the time of detection. The existing studies are based on the assumption that they have all the lifecycle data. But it depends on the time when the news was spread. Fake news has shown adverse effects in a very short time period of dissemination on social media. To avoid this, it is required to detect fake news at an early stage which lacks a sufficient amount of information about the news. Therefore, it is important to analyze the time when the news was disseminated. This paper gives a 'three-phase model' for the detection of fake news: early, middle, and late. Figure 7 graphically represents the variation of fake news detection techniques with time. Table 8 summarizes the three phases of fake news detection in terms of perspective, dataset source, and approach preferred in the literature.

5.1 Early detection

When fake news is published on a news outlet but not yet published on social media. There is a strong need to develop a model to detect fake news early to minimize its social harm. Also, early actions for fake news mitigation and intervention can be taken. It is difficult to control the propagation of fake news once it gets published on social media due to echo chambers and filter bubbles of social media. The methods for early detection focus solely on news content because other social or propagation-based features are unavailable. Zhou et al. proposed a theory-driven fake news detection model by investigating news content at various levels: lexicon-level, syntax-level, semantic-level, and discourse-level [193]. The features have been extracted using standard social and psychological theories and, finally, tested over a supervised machine learning framework. Escalante et al. proposed a novel approach for early detection of threats on social media using profile-based representations (PBRs) [39]. PBR utilizes two tasks, viz., sexual predator detection and aggressive text identification, and conducted experiments using traditional machine learning classifiers. Gereme et al. also introduced a generic model to detect fake news before it flies high by focusing on the content [49]. The experiments conducted showed that deep learning (LSTM-RNN and CNN) models outperformed traditional machine learning classifiers. Essentially, many studies claim that machine learning approaches are incapable of early detection of fake news because they require certain amounts of data to reach decent effectiveness, which takes time to accumulate. However, early detection approaches lead to multiple challenges due to limited information. First, newly appeared events often generate unforeseen knowledge graphs. Second, the content-based models are domain-dependent thus, not generic for all domains. Third, the performance of machine learning classifiers gets deteriorated due to limited available information. Recent research studies provide various ways of dealing with the aforementioned challenges.

5.2 Mid-stage (immediate after posting, before gets viral)

This stage focuses on the content-based features with limited social information to detect fake news immediately after publishing on social media while not yet viral. This is a sound research area currently in this domain. Zhao et al. found that fake news propagates differently from real news at the early stages of disseminating [189]. They explored three features, namely the ratio of layer sizes, the characteristic distance,

Table 6 Popular tools for Fact-checking

Tool	Founded	Product type	Description
The Factual	2016	Mobile app and browser extension	Ranks digital news on a 0-100 scale to evaluate the quality based on source diversity, author expertise, language used.
Logically	2017	Mobile app and browser extension	Automated search assistant feature to assess veracity of information relies on human fact-checkers.
ClaimBuster	2017	Web-based live tool and App	Check factual information relies on NLP and supervised learning.
Grover	2019	AI model	Detect AI generated fake news
Bot sentinel	2018	Platform for Twitter	Categorise Twitter accounts into trustworthy and untrustworthy, and also detect bots.
Sensity AI	2018	Online tool	Identifies the severity of visual threats and useful in deepfakes
Factitious	2017	Gameified format tool	Users (or players) decide on a news whether real or fake and earn points accordingly.
DIRT protocol	2017	Platform	Blockchain verification tool which provides economic incentives to users for improving data accuracy.

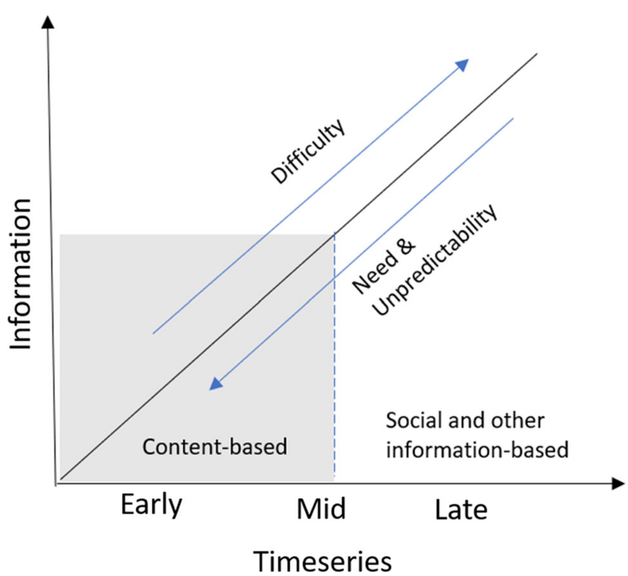


Fig. 7 Variation of fake news detection techniques with time

and the heterogeneity parameter, and tested on support vector machine (SVM) classifier. Since the information available to train a supervised machine learning model is inadequate thus, lead to less accuracy. This limitation has been exploited by Yang Liu et al. by proposing a novel early detection approach. The developed time series approach used propagation paths with convolution, and recurrent neural networks for classification [90]. Yang et al. further proposed a novel FNED model for early detection using text features and users' responses in combination [91]. Compared with existing early fake news detection models, FNED is better because it is content independent, based on less and only relevant features which are required for early detection in the real world. Furthermore, Shu et al. proposed a tri-relationship fake news detection framework (TriFN) by exploring correlations of publisher bias, news stance, and relevant user engagements simultaneously [150]. They observed model performance varying delay time in hours, and the best F1 score was obtained within 48 hours. Nonetheless, the collection of handcrafted features requires rigorous manual efforts. Therefore, Chen et al. proposed a deep attention model using RNN to automatically learn hidden temporal features. It has been observed that users' comments on different contents in different periods of information diffusion [25]. Clearly, the user-oriented features are of great use to early detection of fake news.

5.3 Late detection

The literature shows rich research studies on the detection of fake news after its deep propagation on social media. However, late detection approaches are less helpful for fake news mitigation and intervention as compared to early-stage detection techniques. But research on late detection has given accurate models due to the surplus amount of information available. Early-stage detection models only allow using recent posts within a specific deadline (in hours) since the advent of a particular event, whereas late detection models can use all the available users' posts in the complete time

Table 7 Latest datasets for fake content detection

Content type	Refs.	Dataset	Year	Source	Size	Language/ format	Label	Class distribution	Annotation method	Highest accuracy reported
Text	[177]	LIAR	2017	Politifact	12,836	English	Multiclass (6)	Balanced	Expert-based	95% [166]
	[106]	Credbank	2015	Twitter	60M	English	Multiclass (5)	Balanced	Crowdsourcing-based	48.68% [181]
	[59]	BuzzFeed News	2016	Facebook	2282	English	Multiclass (4)	Imbalanced	Expert-based	92%
	[150]	FakeNewsNet	2017	Buzzfeed Politifact	422	English	Binary	Balanced	Expert-based	93% [140]
	[132]	Fake News Challenge	2017	Fake News Challenge	49,972	English	Multiclass (4)	Imbalanced	Expert-based	99.25% [17]
	[59]	Benjamin Political	2017	Multiple websites	225	English	Multiclass (3)	Balanced	Expert-based (existing dataset)	78% [59]
	[21]	Burfoot Satire News	2009	English Gigaword Corpus	4223	English	Binary	Imbalanced	Expert-based	71% [111]
	[197]	PHEME	2016	Twitter	4842	English & German	Multiclass (3)	Imbalanced	Crowdsourcing-based	86.7% [125]
Image	[115]	IMD2020	2020	Camera models and GANS	74,000	JPEG	Binary	Balanced	Software Manipulations (GANS)	98.81% [70]
	[97]	DEFACTO	2019	MSCOCO	229,000	TIF	Uniary	–	Software Manipulations	88.6% [37]
	[51]	MFC	2019	Random (internet)	16029	RAW, PNG, BMP, JPG, TIF	Binary	Imbalanced	Software Manipulations	82.0% [112]
	[56]	PS-Battles	2018	Photoshopbattles subreddit	1,13170	PNG, JPEG	Binary	Imbalanced	Software Manipulations	88.7% [92]
	[191]	FaceSwap	2017	Online face swapping application	3,685	JPG	Binary	Balanced	Software Manipulations (SwapMe iOS App)	99.8% [191]
Video	[118]	FVC-2018	2018	You-tube Twitter Facebook	6415	Various	Binary	Balanced	Expert-based	0.69 F1-score [118]
	[64]	DeeperForensics-1.0	2020	Manual creation of DeepFake	60000	-	Binary	Imbalanced	Expert-based	64.1% [64]
	[88]	Celeb-DF	2020	Improved synthesis process	6229	MPEG4	Binary	Imbalanced	Software Manipulations	80.58% [85]
	[36]	DFDC-preview	2019	Manual creation of videos	5244	H.264	Binary	Imbalanced	Software Manipulations	82.92% [85]

Binary: Multiclass: Balanced: Imbalanced: Expert-based: Crowdsourcing-based: Software Manipulations:

span of given datasets. Early or mid-stage detection models heavily rely on news content whereas, late detection models additionally can explore network-based cues. Zhou et al. proposed aimed to exploit social network patterns of fake news, which refer to the news content, spreaders of the news, and associations among the spreaders [195]. Nikiforos et al. defined a well-defined fake news detection framework that uses both linguistic and network features [114]. The literature aims to extract novel features to further improve the effectiveness of the existing models. Liao et al. exploited user comments on social posts, which is crucial information but not well studied for fake news detection. They proposed a heterogeneous graph neural network and explainable model to outperform the baseline models [89]. Similarly, Jang et al. proposed a neural network-based fake news classification model using conventional tweet features with an additional new feature, Quote Retweet (Quote RT), introduced by Twitter in 2015. The Quote RT enables users to add a comment while retweeting an existing tweet which leads to tracking the depth of propagation [63].

6 Taxonomies

6.1 Taxonomy based on domain

Fake news has become a substantial social problem with the speedy progression of social media. The dissemination of fake information is not limited to one domain but is present across multiple domains. Fake news is pervasive and has effects across different domains like politics, healthcare, entertainment, terrorism, tourism, and natural catastrophes. Numerous studies have worked on automating fake news detection, which is trained and evaluated using datasets that are limited to a single domain such as politics, entertainment, and healthcare. The majority of the studies have examined political fake news; however, health-related false information is more threatening. Table 9 lists domain-specific studies for the detection of fake news. The techniques proposed in the state of the art have been focused on one domain because the performance of such techniques (machine learning) generally drops if unseen data from different domains appear during training. Features, especially style-based features, are domain variants; thus, features are required to be selected mindfully in order to distinguish fake and true news considering the domain under investigation. State of the art has highlighted this research gap to consider multiple domains. Thus, it is required to develop comprehensive cross-domain approaches for the detection of fake news, although quite a few preceding works have attempted to perform fake news detection using cross-domain datasets (also mentioned few in Table 9). Han et al. proposed a continual learning approach to handle domain agnostic fake news detection. Their approach adopted a graph neural network, which learns different domains sequentially. Nevertheless, it has two limitations: (1) it assumes that the other domain will arrive sequentially, and (2) the other domain is also known, which does not happen in real-world streams. In contrast, the approach proposed by Silva et al. preserved the knowledge about the different domains. They applied the robust, optimized BERT model to decide on informative instances for manual annotation from a large unlabeled corpus. Therefore, existing studies have tried to incorporate datasets from multiple domains to develop a cross-domain fake news detection model. For instance, Castelo et al. [23] have trained the model using the Celebrity dataset (details are given in Table 10) and tested it using the US-Election2016 dataset to evaluate the generalizability of their approach.

6.2 Taxonomy based on features per misinformation type

Particularly, we have described the different types of false information that can be found in OSNs in Sect. 4. In this section, the taxonomy of features per false information type is provided based on the existing literature. Figure 8 presents a two-dimensional view to highlight features per false information by deeply studying different research papers mentioned in Sect. 2. The 2D characterization presents three layers: The innermost layer gives the types of false information, the mid-layer gives the features required for the identification of each type, outermost layer highlights the field per type in which significant work has been done. These identified features per type are required by machine learning algorithms to classify fake news from other types of false information. Though the stated features are not exhaustive and required to be further fine-tuned, however, the features identified per type in this paper can be used as a reference by the new researchers trying to explore this research field. Each stated feature consists of a set of sub-features; for example, the propagation network consists of the number of shares, likes, etc. Also, this figure shows the domain in which the specific misinformation type has been majorly explored in the literature [9].. The prominent features in the state of the art for each false information type have been segregated and analyzed in order to offer a succinct roadmap for future work. For instance, click-baits are attractive headlines with unrelated stories. Thus, similarity analysis is a useful feature to detect clickbait. Similarly, rumors spread differently to reach out maximum audience. Therefore, propagation pattern is an important feature to distinguish rumors from other false information types. Satire contains humorous content; thus, sentiment as a stylistic feature has been used most popularly in the literature. To identify

[9] Web of Science https://www.webofknowledge.com, https://mjl.clarivate.com/home

Table 8 Attributes of the three phases of fake news detection based on literature survey

Three Phases	Time of detection	Perspective	Dataset source	Preferred approach
Early detection	Not yet published on social media	Content-based	News	Deep learning
Middle stage detection	Immediate after posting and before gets viral on social media	Content-based (primarily) + limited social-based features	Social media	Deep learning
Late detection	After deep propagation of news on social media	Social-based features (primarily)	Social media	Deep learning/ Machine learning

Table 9 Domain-specific distribution of literature

Politics	Asubiaro et al. [11], Fairbanks et al. [41], Ribeiro et al. [131], Ajao et al. [5], Shao et al. [148], Patwari et al. [119], Lee et al. [83], Faustini et al. [42], Karimi et al. [68]
Healthcare	Dai et al. [30], Vincent et al. [169], Dhoju et al. [35], Abbasi et al. [1], Hou et al. [60], Kinsora et al. [73]
Terrorism	De et al. [32], Sanchez et al. [142], Cristani et al. [29], Hamdi et al. [54], Kostakos et al. [74], Last et al. [81]
Natural disasters	Wang et al. [175], Gupta et al. [53], Allen et al. [8], Rajdev et al. [126], Krishnan et al. [76], KP et al. [75], Mondal et al. [108]
Tourism and marketing	Kauffmann et al. [69], Cardoso et al. [22], Lee et al. [82], Yoo et al. [184], Chuang et al. [27], Chowdhary et al. [26], Lu et al. [93], Luca et al. [94], Fedeli et al. [44], Fontanarava et al. [46], Juuti et al. [66], Schuckert et al. [143], Banerjee et al. [14], Lappas et al. [80], Mkono et al. [107], Banerjee et al. [13]
Cross-domain	Castelo et al. [23], Perez et al. [120], Saikh et al. [141], Han et al. [55], Silva et al. [153], Gautam et al. [48], Rubin et al. [135], Wang et al. [176]

Table 10 Cross-domain studies

Politifact + Gossipcop	→	Han et al. [55]
Politifact + Gossipcop + CoAID	→	Silva et al. [153]
Celebrity + US-Election2016	→	Castelo et al. [23]
News + Speeches/tweets	→	Wang et al. [176]

Hoaxes, content-based features, as well as sentiment-based features, have been employed [166]. Fake news detection has a wide range of features as per the literature, such as context knowledge and social features [163]. Moreover, fake videos and profiles have also been considered in this domain using user engagement features [164]. To end, features should be selected depending on the type of misinformation.

6.3 Taxonomy based on misinformation data-type

The diffusion of fake news in various formats such as text, image, audio, video, and links on online information platforms is a fast-growing problem. The multimedia content, including images and videos, allure users more than text. So far, most of the ongoing research has focused on one data modality, and limited work has been done considering both textual and multimedia content on social media. Thus, one of the key research challenges in fake news detection is multimodal data verification. This taxonomy provides details of existing studies based on the type of data.

1. Textual data: Textual content generally studies linguistics cues. A rich literature exists to develop fake news detection models considering only text data using textual and user metadata features. The majority of existing studies done at the text level have exploited style of writing as a prominent feature because style-based features capture authenticity as well as intention. Popular style-based features like linguistic features such as n-grams [2], psycholinguistic features using LIWC, number of punctuations, stopwords, readability scores (e.g., number of complex words, long words, syllables, characters) [120], syntax and dictionary-based features [121]. Psychological features such as sentiment, emotion are strong differentiating factors between fake and real content [59] [4]. Siering et al. proposed a framework based on the verbal cues of the content (e.g., average sentence length, subjectivity, PoS) to know the deception process, the psychology of fake spreaders, and type of cues [152]. Apart from textual handcrafted features, the literature highlights various studies based on latent features for news text embedding. Such embeddings are processed at the word level,

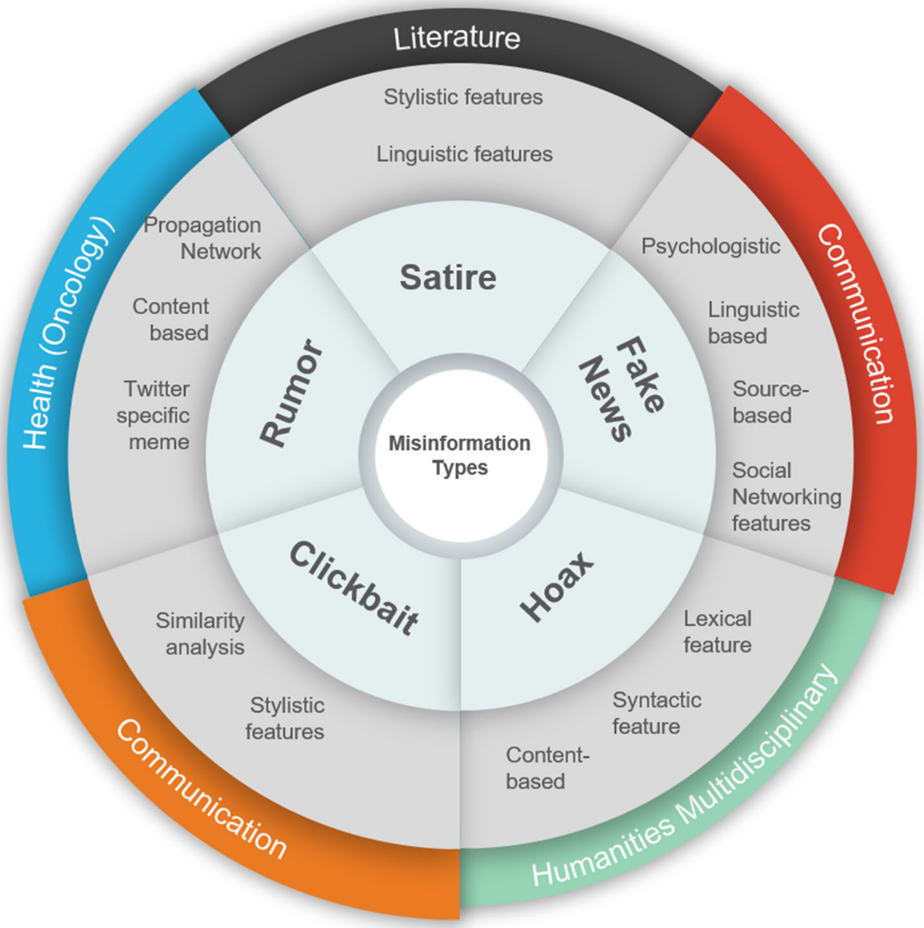


Fig. 8 2D characterization feature sets per false information type

sentence level or document level [105] to obtain vectors representing text which can further be given to machine learning [50] [71] or deep learning classifiers [67] [157].

2. Image: Image is a part of multimedia content; however, it has a standalone medium to be a news source. Image forensics has been long used to evaluate the authenticity of images by checking whether a digital image has been manipulated. Image modification using image-to-image translation techniques such as generative adversarial networks (GANs) is done realistically. Marra et al. studied and compared the performance of various image forgery detectors [99]. Also, Hsu et al. employed GANs to produce fake-real image pairs and then proposed a deep learning approach to detect fake images using contrastive loss [61]. On the other hand, Vishwakarma et al. developed a reverse algorithm to check the veracity of image text by searching it on the web and evaluating the credibility of content using reality parameter Rp [171]. Some of the existing studies have considered different formats of data in an integrated form. Dun Li et al. integrated text, image, propagation, and user-based features to improve the performance of the fake news detection model [84]. Yang et al. identified explicit and latent features of both text and images and proposed a model named TI-CNN (Text and Image information based Convolutional Neural Network) [183]. Boididou et al. also used textual features along with image forensic to check the authenticity of multimedia content on Twitter [19].

3. Video: A rich related literature focuses on tampering detection and image/video forensics algorithms. The same methods cannot be used for image and video detection due to the deterioration of the frame data after video compression. Papadopoulou et al. [118] presented a corpus of real and fake user-generated videos and a classification model based on video metadata features. They observed that fake videos are of shorter duration, fake videos tend to be posted by younger Twitter accounts, and the text that accompanies fake videos have distinctive linguistic qualities. Clickbait is a type of false content as described in Sect. 2.4. For this, Varshney et al. [168] proposed a clickbait video detector (CVD) scheme based on three latent features sets, namely video content (e.g., title-video similarity), user profile (e.g., registration age), and human consensus (e.g., user comments). Video content features

have been extracted by mining audio from the videos and, later on, transformed into text data to further extract text-based features. Finally, all the features have been integrated and given to the classification model. Recently, due to the admirable generation capability of generative adversarial networks (GANs), it has been used for image generation and video predictions. Dong et al. [38] used the idea of GANs for video anomaly detection and executed few experiments to prove the efficacy of their approach. Furthermore, GANs have led to the generation of synthetic videos that closely resemble real videos known as deepfakes. Deepfakes is a novel form of fake content originated by combining deep learning and fake content. Deepfakes are extremely realistic, thus, hard to detect. Vizoso et al. [172] disclosed the effect of this new form of fake news (i.e., deepfakes) on different popular news media and social media platforms. Several studies to detect deepfakes have been presented in state-of-the-art but still limited in terms of setting a benchmark for validating different detection methods. One of the reasons is the absence of a gold standard dataset in this domain. Deepfakes detection methods are based on two approaches, namely temporal features across video frames and visual artefacts within the video frame. Temporal feature-based methods have mostly used deep recurrent network methods for deepfake videos detection. Sabir et al. [137] extracted spatiotemporal features of videos, while [133] proposed a facial manipulation detection method and tested it on FaceForensics++ data set. On the other hand, methods based on visual artefacts break the video into frames and extract features for the individual frame to distinguish between fake and real videos. Generally, deepfake videos are of low resolution; hence, CNN models such as VGGs and ResNets [87] can be used to detect the resolution inconsistencies. Yang et al. [182] proposed an SVM classifier using 68 facial landmark features to classify deepfakes and real images or videos. Despite the strong ongoing research and several forensic tools, there is still a need for new and timely solutions in multimedia forensics.

4. Multimodal: However, the aforementioned uni-modal techniques have offered promising results, but the unceremonious behavior of online social media data has always been a hurdle in data extraction. Thus, several studies have been working with multiple modalities, e.g., text, images, and videos. Wang et al. [178] developed an end-to-end model named Event adversarial neural network (EANN) for multimodal fake news detection. Two different techniques followed for text and image, i.e., word embeddings using CNN and VGG-19 on ImageNet, respectively, and finally, both fused into a neural network classifier. On the other hand, Zhou et al. [194] developed a multimodal termed SAFE to find the relationship between textual and visual features. They observed that fake content creators use irrelevant features to allure users. Further, Khattar et al. developed a model inspired by [178] and called it multimodal variational autoencoder for fake news detection (MVAE). A bidirectional LSTM was used to learn text vectors, while the same VGG-19 was used to learn image features. Meel et al. also used VGG-19 and Bi-LSTM to analyze veracity in multimodal data [101]. For simplicity and generalizability of the systems, Singhal et al. [156] developed SpotFake, a multimodal fake news detection model using the BERT language model to extract text features, and image features were again extracted using VGG-19 pre-trained on ImageNet dataset.

6.4 Taxonomy based on platform

Table 11 categorizes the research papers according to the platforms they study. This is an important categorization because data collection and features are two vital steps in machine learning algorithms and these are platform-specific. The table also shows few studies which have extracted data from multiple platforms and then merged them to develop a multi-platform dataset. Furthermore, there are limited research works which have proposed frameworks to fetch data from multiple platforms [162] [165] [144].

To end, the performance of a supervised machine learning algorithm depends on features, domain, data format, and platform. Thus, these taxonomies are useful to make sense of the existing work in different categories.

7 Bibliometric analysis

This section visualizes the bibliometric analysis for this study based on 'Incites' (https://incites.clarivate.com). It is the second major contribution of the paper apart from the comprehensive overview of the literature in multiple directions. The figures and analysis presented in this section are useful to get complete knowledge about this research area, such as funding agencies and top organizations working in this area, the geographic areas essentially publishing papers in this domain, foremost journals, as per the Web of Science records. Moreover, this rigorous bibliometric analysis of scientific data will help established and emerging researchers to uncover journal trends and impact, co-words networks to understand the thematic structure, etc. Therefore, this section is useful to gain a comprehensive one-stop overview of the work done in this domain.

7.1 Statistics according to quartile

Based on Impact Factor (IF) data, the Journal Citation Reports published by Thomson Reuters (https://incites.

Table 11 Taxonomy based on platform

Social Platforms	Research Papers
Twitter	Monti et al. [109], Ajao et al. [5], Bessi et al. [16], Gupta et al. [53], Jin et al. [65], Kim et al. [72],Helmstetter et al. [57], Nied et al. [113], Ruchansky et al. [136]
Facebook	Preston et al. [122], Tachhini et al. [160], Del et al. [34], Silverman et al. [154]
Review platforms	Akoglu et al. [6], Hooi et al. [58], Kumar et al. [77], Barbado et al. [15], Mukherjee et al. [110], Shan et al. [146]
Sina Weibo	Guo et al. [52], Kim et al. [72], Zhou et al. [192], Ruchansky et al. [136], Wu et al. [180]
Multi-platform	Twitter+Sina Weibo+News articles: Faustini et al. [42], Reddit+Twitter+4chan: Zannettou et al. [186]
Fake news articles	Horne et al. [59], Silverman et al. [154], Rubin et al. [135], Perez et al. [120]
Wikipedia	Kumar et al. [79], Saez et al. [138], Solorio et al. [159]
Fact checking websites	Shu et al. [150], Shahi et al. [145]
Crowdsourcing	Perez et al. [120]

clarivate.com) provide yearly rankings of science and social science journals, in the subject categories relevant for the journal (in fact, there may be more than one). Quartile rankings are therefore derived for each journal in each of its subject categories according to which quartile of the IF distribution the journal occupies for that subject category. Q1 denotes the top 25% of the IF distribution, Q2 for the middle-high position (between top 50% and top 25%), Q3 middle-low position (top 75% to top 50%), and Q4 the lowest position (bottom 25% of the IF distribution). In this analysis, the articles have been explored using keywords like Fake news + Misinformation + Rumors + Hoax + Satire + Clickbait. A total of 8016 articles have been extracted with different subject categories, and the maximum cited articles fall in the subject category, *Oncology*. However, the articles under Oncology are not of much relevance as far as this review analysis is concerned. Since the research on misinformation was highlighted during the United States Presidential election in 2016. Therefore, communication, as shown in Fig. 9, became a hot research area for the past decade for the keywords, and maximum papers were published in 2019. Figure 10 shows the temporal trends of the Journals in this domain.

Table 12 and Table 13 provide the journal names and their quartiles. Impact factors of different quartile journals are provided in Fig. 11.

7.2 Statistics based on different geographic locations

A total of 293 journals were obtained using keywords described in Sect. 7.1, covering areas like communication, computer science–artificial intelligence, interdisciplinary applications, and many other related research areas. Since the topic of this review is not limited to a particular domain and became active after 2016 specifically, a total of 252 relevant Journals have been analyzed. It has been observed that most of the Web of science documents are from the USA, as shown in Fig. 12. Furthermore, the location of researchers who have published a high number of papers in this domain is also the USA. Table 14 shows the top collaborating organizations have been working most actively in this area. Moreover, most frequent collaborating academic organizations have also been analyzed for refined view in Fig. 13. Figure 14 shows the most active funding agencies based on the number of documents in Web of science in this domain worldwide. For extended analysis, network visualization map and density visualization map has been produced using 'VOSviewer' (https://www.vosviewer.com/). Network Visualization consists of items (circles and labels) and links (relation between the items). Items are grouped into clusters specified by the different colors assigned to the items. Figure 15 shows the network visualization of keywords of the reviewed articles, and the link between them shows their co-occurrences. Figure 15 shows the keyword 'fake news' has the highest occurrences, and it is highly co-occurred with 'misinformation,' 'social media,' and 'twitter.' Figure 16 shows the network visualization of the terms taken from the abstract and title of the reviewed articles; links indicate the number of articles in which two terms occur together. In this, the term 'rumor' has the highest occurrences, and it strongly co-occurred with 'network,' 'model,' and 'time.'

8 Potential future directions

The literature highlights several studies on false information, but the problem is still unresolved and still a hot topic in research. We have outlined some key areas required to be addressed in the future.

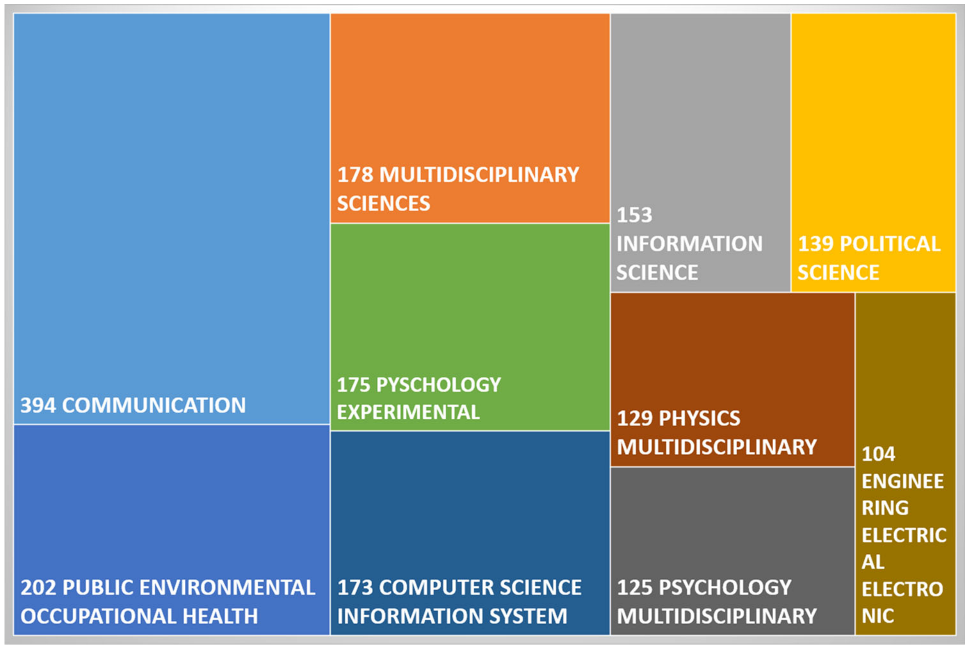


Fig. 9 Articles count per research category

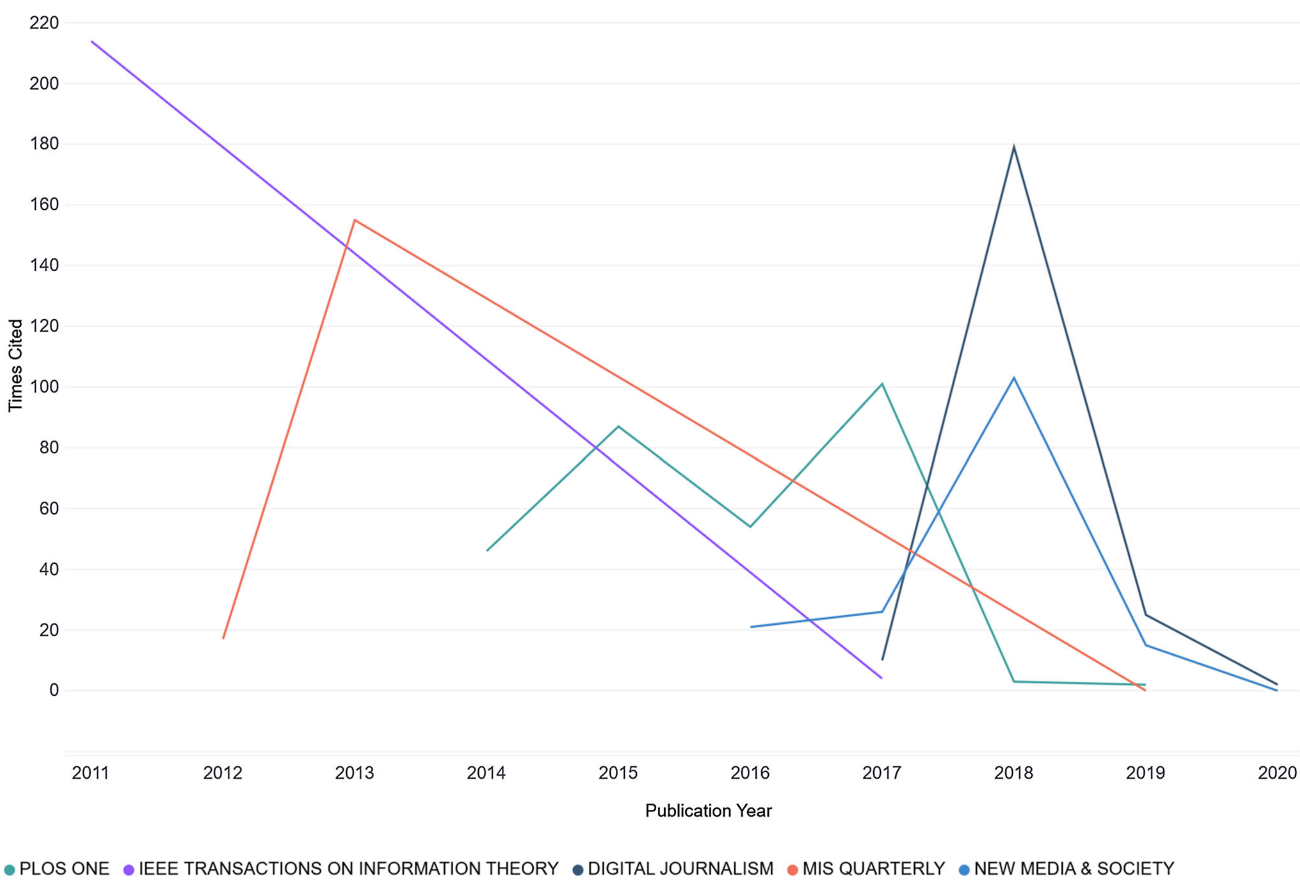


Fig. 10 Trends Graph of Journals

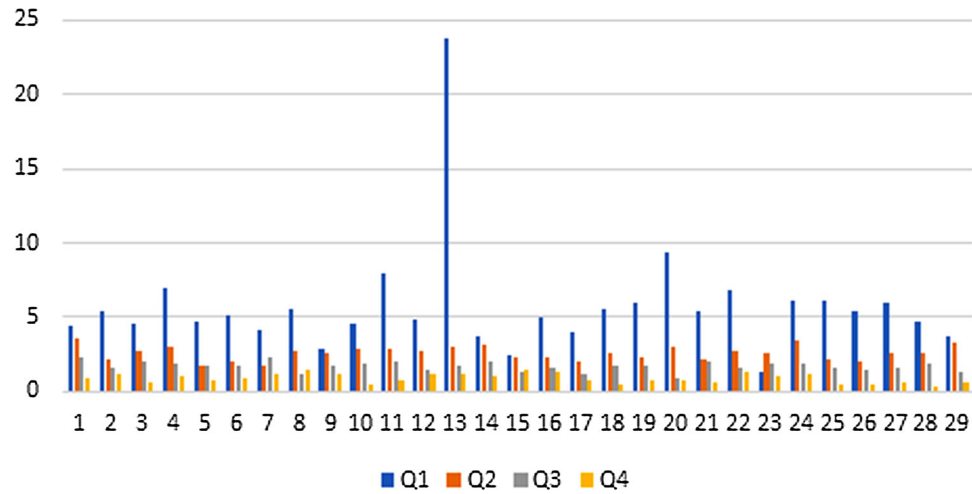


Fig. 11 Impact factor of different journals based on quartile

Table 12 Journal name and ranking for quartiles Q1 & Q2

Q1	Q2
Digital journalism	Information systems frontiers
MIS quarterly	Mass communication quarterly
New media & society	Social science computer review
Communications of the acm	International journal of medical informatics
Internet research	Journal of documentation
Political communication	Media culture & society
Journal of the american medical informatics association	Mass communication and society
Journal of network and computer applications	Acm transactions on intelligent systems and technology
Social media + society	Data mining and knowledge discovery
Information communication & society	Acm transactions on information systems
Acm computing surveys	Proceedings of the vldb endowment
Information processing & Management	plos one
Ieee communications surveys and tutorials	Ieee transactions on information theory
Ieee access	Computer networks
European journal of communication	Journalism studies
Ieee transactions on knowledge and data engineering	Science communication
Organization studies	Health communication
Bioinformatics	Ieee transactions on parallel and distributed systems
Information sciences	Journal of parallel and distributed computing
Proceedings of the national academy of sciences of the united states of america	Soft computing
Expert systems with applications	International journal of systems science
Ieee transactions on dependable and secure computing	Peer-to-peer networking and applications
Ieee transactions on systems man cybernetics-systems	Swarm intelligence
Future generation computer systems-the international journal of escience	Cluster computing-the journal of networks software tools and applications
Ieee transactions on multimedia	Studies in informatics and control
Journal of computer-mediated communication	Central european journal of operations research
Ieee transactions on information forensics and security	Symmetry-basel
Computers environment and urban systems	European journal of innovation management
Communication research	Applied intelligence

Table 13 Journal name and ranking for quartiles Q3 & Q4

Q3	Q4
European management journal	Discrete event dynamic systems-theory and applications
Profesional de la informacion	International journal of communication
Acm transactions on knowledge discovery from data	Information processing letters
Acm sigcomm computer communication review	Javnost-the public
Online information review	Asian journal of communication
Journal of the ACM	Journal of combinatorial optimization
Journal of nursing management	Acm transactions on the web
Communication culture & critique	European management review
Kybernetes	Journal of organizational computing and electronic commerce
Journal of intelligent & fuzzy systems	African journalism studies
Journal of real-time image processing	Acm transactions on algorithms
Concurrency and computation-practice & experience	International journal of distributed sensor networks
International journal of human-computer interaction	Applied stochastic models in business and industry
Journal of contingencies and crisis management	Iet information security
Critical discourse studies	International journal of sensor networks
International Journal of Information Security	International journal of uncertainty fuzziness and knowledge-based systems
Journal of experimental & theoretical artificial intelligence	Ieee Latin America transactions
Baltic Journal of Management	International journal of foundations of computer science
International journal on semantic web and information systems	Science technology and society
Information and computation	Sigmod record
Personal and ubiquitous computing	Intelligent data analysis
Expert systems	Intelligent automation and soft computing
Scandinavian journal of management	Simulation-transactions of the society for modelling and simulation international
Iet biometrics	Applied artificial intelligence
Journalism practice	Journal of information science and engineering
Media and communication	Computing and informatics
Computational and mathematical organization theory	Ksii transactions on internet and information systems
Cognitive systems research	Journal of web engineering
Mathematical and computer modelling of dynamical systems	Turkish journal of electrical engineering and computer sciences

- Lack of gold standard dataset. A comprehensive fake news dataset demands a cross-domain, cross-language, or cross-topic analysis, but it has been less explored in the current literature. Such datasets could empower detection similar to real-time research.
- Mostly, studies address authenticity as a measure to evaluate false news, and very few studies focus on the intent of false information. Expert data annotation is one way to capture the intention behind a piece of news. Moreover, mindfully selecting features considering well-established theories of social sciences is another approach to reflect the intention.
- Detection of newly emerging fake news using historical data is a need of the hour, i.e., early prediction of fake news.
- In addition to supervised learning, semi-supervised and unsupervised learning can also play an important role.
- Not sufficient research has been done on designing news feed algorithms robust to fake news propagation. Thus, there is a need to eliminate echo chambers and biased search engines in order to combat fake news.
- A lot of work has been done to detect false information on online sources. Nevertheless, mitigation of false information is still not covered yet.

9 Analysis and conclusion

The identification of false information is an emerging area due to the rise of social media. To automate the process of identification, researchers have proposed several machine-learning-based solutions. However, machine-learning techniques require a large training dataset, a relevant set of features, and the selection of an appropriate classifier based

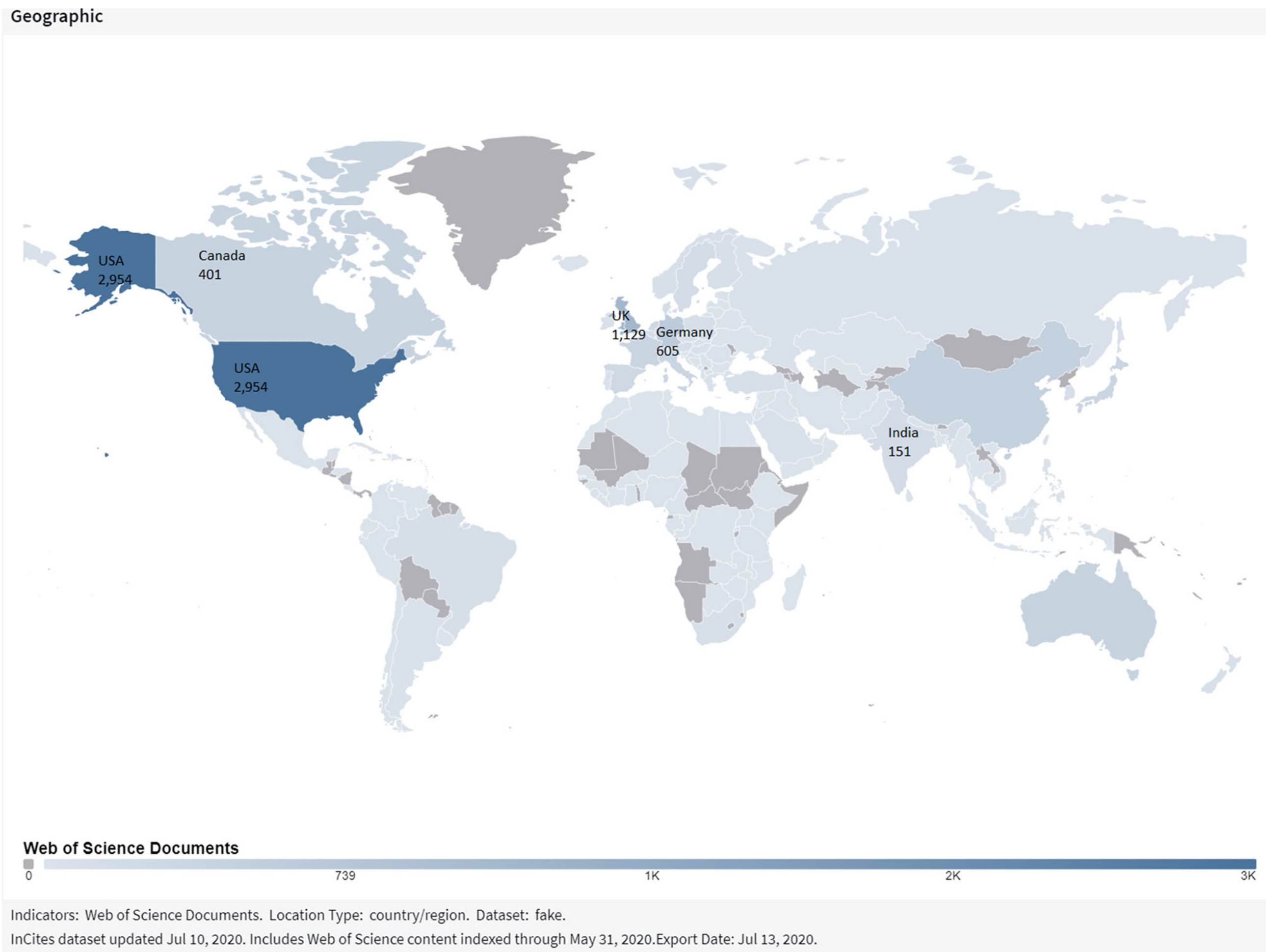


Fig. 12 Number of Web of Science documents based on geographic locations

on the given data [103]. Therefore, this survey covers different aspects that affect the performance of the machine-learning technique. The survey is done in two segments. The first segment offers a comprehensive overview of false information detection, and in the second segment, bibliometric analysis is provided.

The first segment of the survey has mainly focused on different elements required to be considered before designing a methodology. We presented the typology of false information with clear definitions and real-life examples. The idea is to compare the existing work based on the type of false information. Hence, the researchers can design their methodology based on the type of information in order to understand the harmfulness of the news and obtain relevant features for its classification. Another important element is to check the time of detection. Current studies are based on an assumption that they have all the lifecycle data. But it depends on the time when the news was spread. The content and network-based features vary with the time at which news was spread and the time of detection. Our survey highlights this gap by presenting a novel three-phase model with early, mid-, and late stages.

A combined element of four taxonomies has also been presented in the paper. The dissemination of fake information is not limited to one domain but is present across multiple domains. The effect of fake news varies in domains like politics, healthcare, entertainment, terrorism, tourism, and natural catastrophes. For example, fake news related to tourism is less harmful than terrorism. Political fake news spreads faster than other fake news [173]; thus, the propagation features for the political domain are stronger as compared to fake news in other domains. Also, the performance of a supervised machine learning algorithm degrades when the model is trained on one domain and tested on an unseen domain. Thus, the paper presents a taxonomy of existing studies based on a particular domain and also cross-domain scenarios. This will help the researchers to select appropriate features and algorithms according to the given domain. Moreover, fake news is spread in different data formats, i.e., textual as well as multimedia. We have segre-

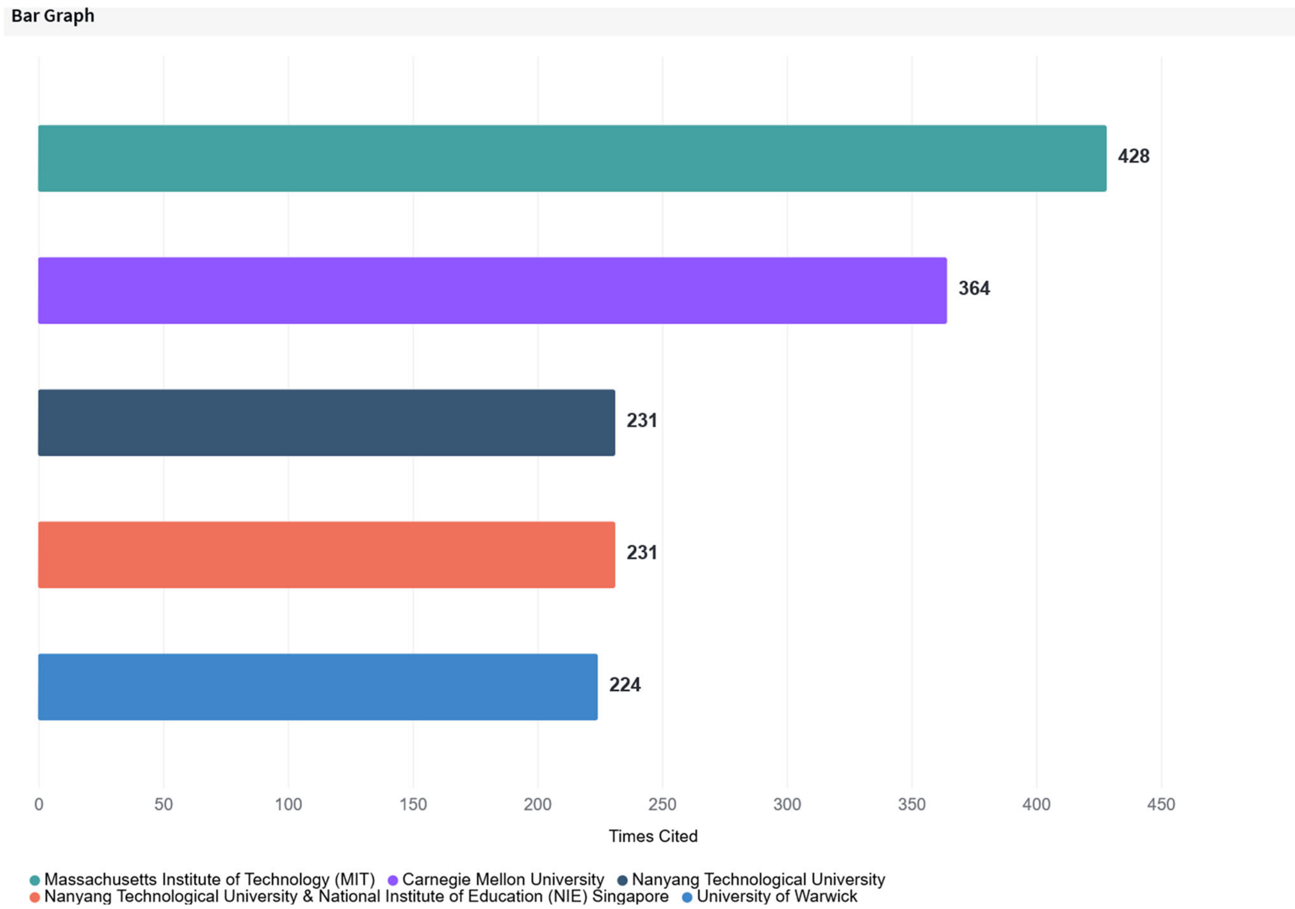
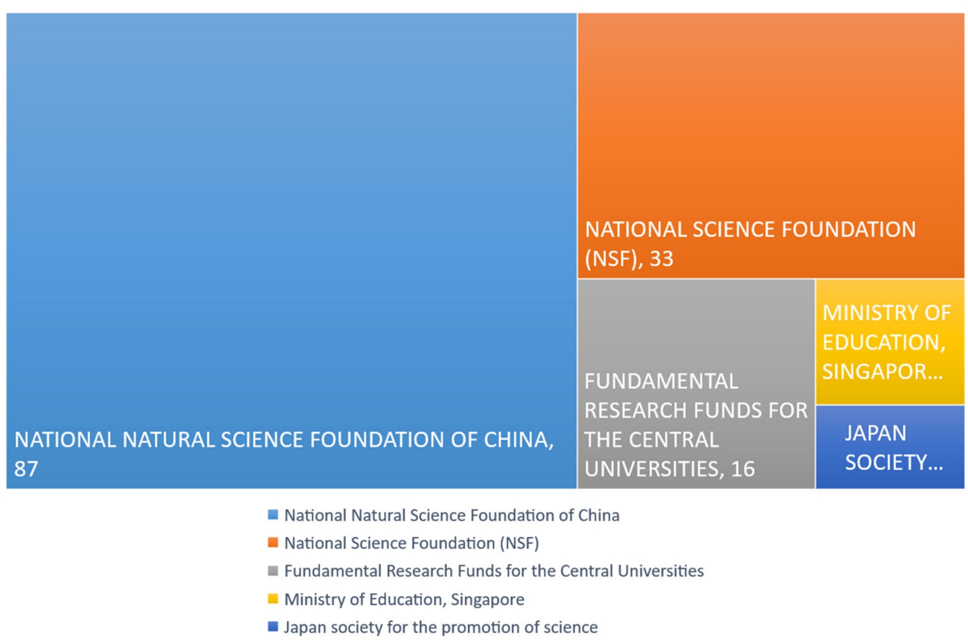


Fig. 13 Organizations based on the number of times cited

Fig. 14 Funding agencies worldwide



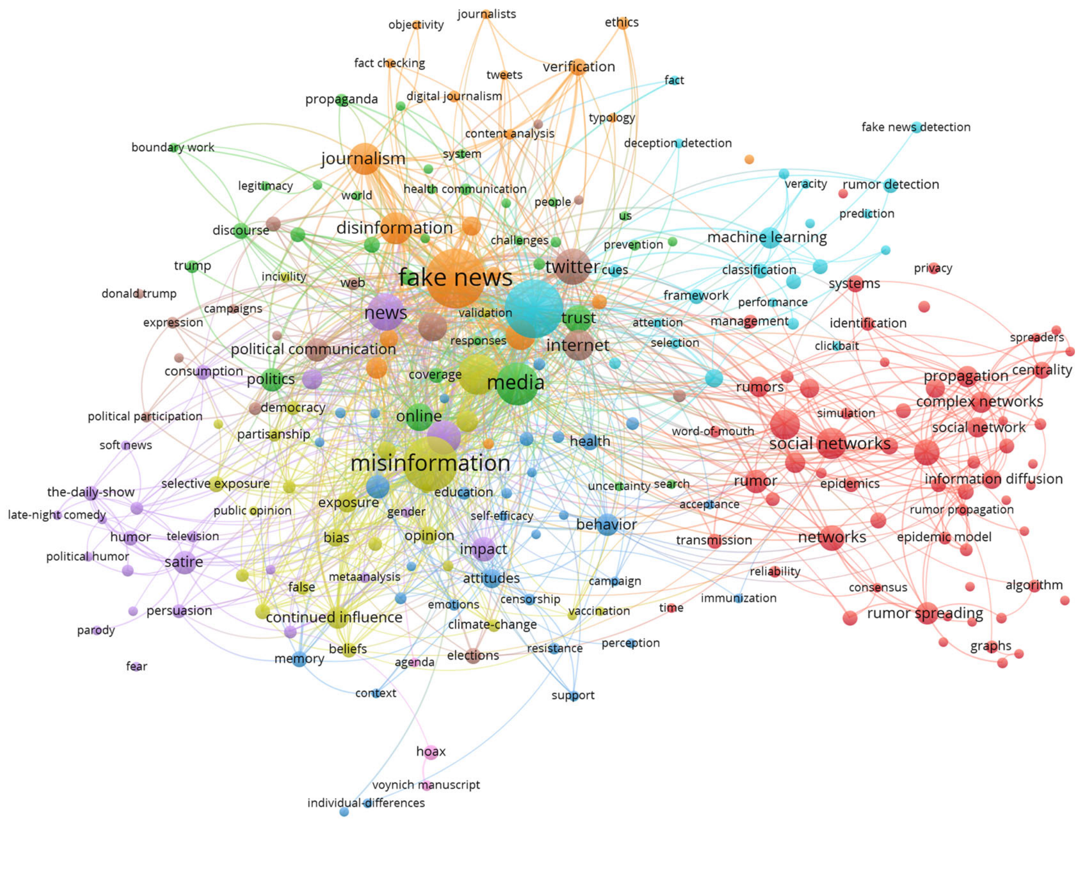


Fig. 15 Network visualization of terms

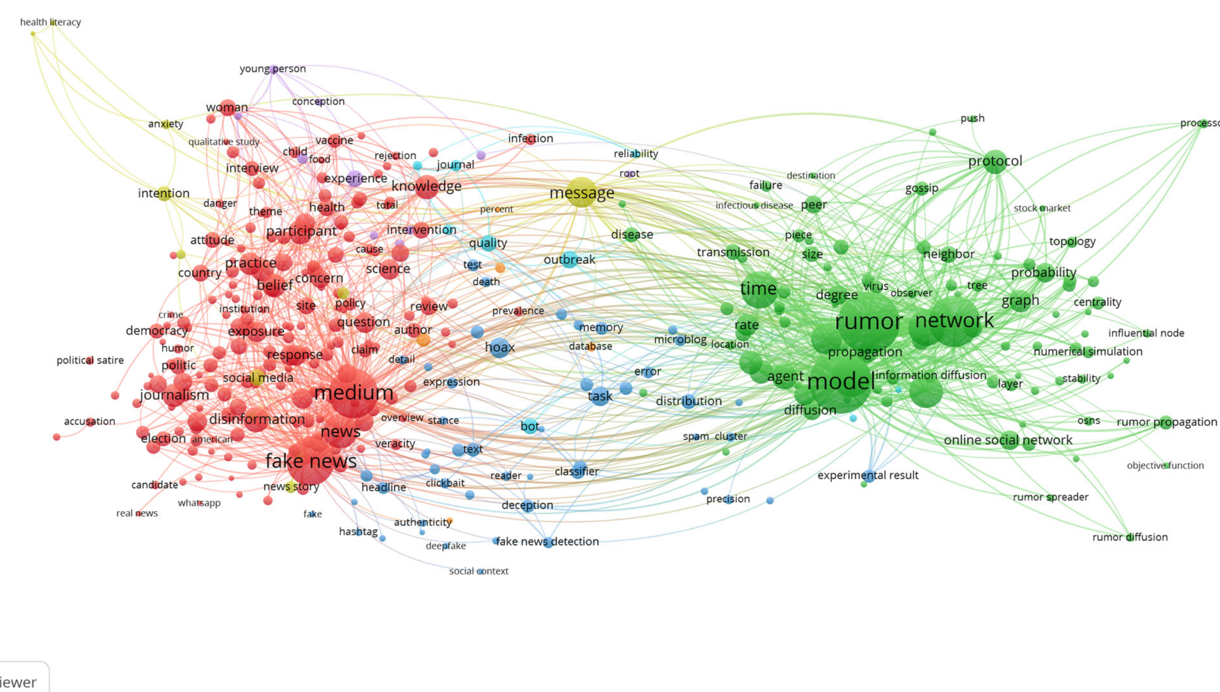


Fig. 16 Network visualization of keywords

Table 14 Top collaborating academic organizations

Organization	No. of papers in related areas
Harvard University	37,971
University of California system	15,816
Broad Institute	12,911
United States department of energy (DOE)	11,208
Massachusetts general hospital	10,449
Brigham and women's hospital	6968
Centre national de la recherche scientific	6801
Lincoln Laboratory	5836
University of California Berkeley	5812
Harvard medical school	5660
Helmholtz association	5298
California institute of technology	5297
Howard Hughes medical institute	5229
Stanford University	5221
University of Chicago	5199
Boston University	5176
University of Texas system	4969
Princeton University	4939
University of California Los Angeles	4745
University of California San Diego	4721

gated the current studies based on unimodal and multimodal research. The survey presents another taxonomy based on platforms such as Twitter, Facebook, and WhatsApp. It is useful to understand how fake news spread on different platforms. The data collection, features, algorithms, and even the impact of fake news, etc., vary with the platform on which the fake news is spreading.

The second segment of our survey provides a bibliometric analysis. This will be useful to the readers in finding the foremost journals, publication houses, organizations, funding agencies, and keywords related to this area. The demographic spread may also assist in interdisciplinary research. Readers require to know such information to comprehend the domain completely apart from the literature review.

In addition, the survey presents a list of actors involved in spreading fake news and actions taken by service providers. This is useful to understand the destructiveness of the content in order to take appropriate actions. An extensive comparison of available datasets based on language, format, labels, annotation technique, etc. Such information can significantly contribute to selecting proper datasets for their research, which can be multilingual or focus on low-resource languages. It can also help to select evaluation metrics and evaluate the annotation quality.

Therefore, the survey aims to provide multiple directions required to be considered for designing an effective machine-learning-based solution. The elements presented are not solely accountable for the detection of false information, but are an important aspect to plan a coherent methodology.

In future, more research is needed to specify how fake news spread across different social media platforms. Finding and categorizing social media will also be an important step before an upcoming network visualization that could possibly point toward representing the outcomes visually.

Author Contributions SR: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Visualization. DB: Writing - Review and Editing, Supervision.

Funding The authors did not receive support from any organization for the submitted work.

Availability of data and material Not applicable

Code Availability Not applicable.

Declarations

Conflicts of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Consent to participate Not applicable.

Consent for publication Not applicable.

Ethics approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Abbasi, A., Zahedi, F.M., Kaza, S.: Detecting fake medical web sites using recursive trust labeling. ACM Trans. Inf. Syst. (TOIS) **30**(4), 1–36 (2012)
2. Ahmed, H., Traore, I., Saad, S.: 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, pp. 127–138. Springer (2017)
3. Ahmed, S., Hinkelmann, K., Corradini, F.: Combining machine learning with knowledge engineering to detect fake news in social networks-a survey. In: Proceedings of the AAAI 2019 Spring Symposium, vol. 12 (2019)
4. Ahuja, R., Bansal, S., Prakash, S., Venkataraman, K., Banga, A.: Comparative study of different sarcasm detection algorithms based on behavioral approach. Procedia Comput. Sci. **143**, 411–418 (2018)
5. Ajao, O., Bhowmik, D., Zargari, S.: Fake news identification on twitter with hybrid cnn and rnn models. In: Proceedings of the 9th International Conference on Social Media and Society, pp. 226–230 (2018)
6. Akoglu, L., Chandy, R., Faloutsos, C.: Opinion fraud detection in online reviews by network effects. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 7 (2013)
7. Alkhodair, S.A., Ding, S.H., Fung, B.C., Liu, J.: Detecting breaking news rumors of emerging topics in social media. Inf. Process. Manag. **57**(2), 102018 (2020)
8. Allen, D.E., McAleer, M.: Fake news and indifference to scientific fact: president trump' confused tweets on global warming, climate change and weather. Scientometrics **117**(1), 625–629 (2018)
9. Almaatouq, A., Shmueli, E., Nouh, M., Alabdulkareem, A., Singh, V.K., Alsaleh, M., Alarifi, A., Alfaris, A., et al.: If it looks like a spammer and behaves like a spammer, it must be a spammer: analysis and detection of microblogging spam accounts. Int. J. Inf. Secur. **15**(5), 475–491 (2016)
10. Alzanin, S.M., Azmi, A.M.: Rumor detection in Arabic tweets using semi-supervised and unsupervised expectation-maximization. Knowl.-Based Syst. **185**, 104945 (2019)
11. Asubiaro, T.V., Rubin, V.L.: Comparing features of fabricated and legitimate political news in digital environments (2016–2017). Proc. Assoc. Inf. Sci. Technol. **55**(1), 747–750 (2018)
12. Baines, D., Elliott, R., et al.: Defining misinformation, disinformation and malinformation: An urgent need for clarity during the covid-19 infodemic. Discussion Papers **20** (2020)
13. Banerjee, S., Chua, A.Y.: Understanding the process of writing fake online reviews. In: Ninth International Conference on Digital Information Management (ICDIM 2014), pp. 68–73. IEEE (2014)
14. Banerjee, S., Chua, A.Y., Kim, J.J.: Using supervised learning to classify authentic and fake online reviews. In: Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication, pp. 1–7 (2015)
15. Barbado, R., Araque, O., Iglesias, C.A.: A framework for fake review detection in online consumer electronics retailers. Inf. Process. Manag. **56**(4), 1234–1244 (2019)
16. Bessi, A., Ferrara, E.: Social bots distort the 2016 US presidential election online discussion. First Monday **21**(11–7), 14 (2016)
17. Bhatt, G., Sharma, A., Sharma, S., Nagpal, A., Raman, B., Mittal, A.: On the benefit of combining neural, statistical and external features for fake news identification. arXiv preprint arXiv:1712.03935 (2017)
18. Biyani, P., Tsioutsiouliklis, K., Blackmer, J.: " 8 amazing secrets for getting more clicks": detecting clickbaits in news streams using article informality. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 30 (2016)
19. Boididou, C., Middleton, S.E., Jin, Z., Papadopoulos, S., Dang-Nguyen, D.T., Boato, G., Kompatsiaris, Y.: Verifying information with multimedia content on twitter. Multimed. Tools. Appl. **77**(12), 15545–15571 (2018)
20. Bondielli, A., Marcelloni, F.: A survey on fake news and rumour detection techniques. Inf. Sci. **497**, 38–55 (2019)
21. Burfoot, C., Baldwin, T.: Automatic satire detection: are you having a laugh? In: Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, pp. 161–164 (2009)
22. Cardoso, E.F., Silva, R.M., Almeida, T.A.: Towards automatic filtering of fake reviews. Neurocomputing **309**, 106–116 (2018)
23. Castelo, S., Almeida, T., Elghafari, A., Santos, A., Pham, K., Nakamura, E., Freire, J.: A topic-agnostic approach for identifying fake news pages. In: Companion Proceedings of the 2019 World Wide Web Conference, pp. 975–980 (2019)
24. Chafkin, M.: How snapchat has kept itself free of fake news. Bloomberg Businessweek **26** (2017)
25. Chen, T., Li, X., Yin, H., Zhang, J.: Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 40–52. Springer (2018)
26. Chowdhary, N.S., Pandit, A.A.: Fake review detection using classification. Int. J. Comput. Appl. **180**(50), 16–21 (2018)
27. Chuang, Y.C., Li, Y.M.: Detecting spam reviews for improving house sharing recommendation. In: 2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI), pp. 91–94. IEEE (2019)
28. Crawford, K., Gillespie, T.: What is a flag for? Social media reporting tools and the vocabulary of complaint. New Media Soc. **18**(3), 410–428 (2016)
29. Cristani, M., Burato, E., Santacá, K., Tomazzoli, C.: The spider-man behavior protocol: exploring both public and dark social networks for fake identity detection in terrorism informatics. KDWeb **1489**, 77–88 (2015)
30. Dai, E., Sun, Y., Wang, S.: Ginger cannot cure cancer: Battling fake health news with a comprehensive data repository. arXiv preprint arXiv:2002.00837 (2020)
31. Daoud, D.M., El-Seoud, M.: An effective approach for clickbait detection based on supervised machine learning technique. Int. J. Online Biomed. Eng. **15**(3), 21–32 (2019)
32. De Smedt, T., De Pauw, G., Van Ostaeyen, P.: Automatic detection of online jihadist hate speech. arXiv preprint arXiv:1803.04596 (2018)
33. Dearden, E., Baron, A.: Fool's errand: looking at April fools hoaxes as disinformation through the lens of deception and humour (2019)
34. Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H.E., Quattrociocchi, W.: The spreading of misinformation online. Proc. Natl. Acad. Sci. **113**(3), 554–559 (2016)
35. Dhoju, S., Main Uddin Rony, M., Ashad Kabir, M., Hassan, N.: Differences in health news from reliable and unreliable media. In: Companion Proceedings of The 2019 World Wide Web Conference, pp. 981–987 (2019)
36. Dolhansky, B., Howes, R., Pflaum, B., Baram, N., Ferrer, C.C.: The deepfake detection challenge (dfdc) preview dataset. arXiv preprint arXiv:1910.08854 (2019)
37. Dong, C., Chen, X., Hu, R., Cao, J., Li, X.: Mvss-net: Multi-view multi-scale supervised networks for image manipulation detection. arXiv preprint arXiv:2112.08935 (2021)
38. Dong, F., Zhang, Y., Nie, X.: Dual discriminator generative adversarial network for video anomaly detection. IEEE Access **8**, 88170–88176 (2020)
39. Escalante, H.J., Villatoro-Tello, E., Garza, S.E., López-Monroy, A.P., Montes-y Gómez, M., Villaseñor-Pineda, L.: Early detection

of deception and aggressiveness using profile-based representations. Expert Syst. Appl. **89**, 99–111 (2017)
40. Faghani, M.R., Nguyen, U.T.: Mobile botnets meet social networks: design and analysis of a new type of botnet. Int. J. Inf. Secur. **18**(4), 423–449 (2019)
41. Fairbanks, J., Fitch, N., Knauf, N., Briscoe, E.: Credibility assessment in the news: Do we need to read. In: Proc. of the MIS2 Workshop held in Conjuction with 11th Int'l Conf. on Web Search and Data Mining, pp. 799–800 (2018)
42. Faustini, P.H.A., Covões, T.F.: Fake news detection in multiple platforms and languages. Expert Syst. Appl. **158**, 113503 (2020)
43. Fauzi, A., Setiawan, E., Baizal, Z.: Hoax news detection on twitter using term frequency inverse document frequency and support vector machine method. In: Journal of Physics: Conference Series, vol. 1192, p. 012025. IOP Publishing (2019)
44. Fedeli, G.: Fake news" meets tourism: a proposed research agenda. Ann. Tour. Res. **126**(1), 21 (2019)
45. Figueira, Á., Oliveira, L.: The current state of fake news: challenges and opportunities. Procedia Comput. Sci. **121**, 817–825 (2017)
46. Fontanarava, J., Pasi, G., Viviani, M.: Feature analysis for fake review detection through supervised classification. In: 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 658–666. IEEE (2017)
47. Franz, D., Marsh, H.E., Chen, J.I., Teo, A.R.: Using facebook for qualitative research: a brief primer. J. Med. Internet Res. **21**(8), e13544 (2019)
48. Gautam, A., Jerripothula, K.R.: Sgg: Spinbot, grammarly and glove based fake news detection. In: 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM), pp. 174–182. IEEE (2020)
49. Gereme, F.B., Zhu, W.: Early detection of fake news" before it flies high". In: Proceedings of the 2nd International Conference on Big Data Technologies, pp. 142–148 (2019)
50. Gravanis, G., Vakali, A., Diamantaras, K., Karadais, P.: Behind the cues: a benchmarking study for fake news detection. Expert Syst. Appl. **128**, 201–213 (2019)
51. Guan, H., Kozak, M., Robertson, E., Lee, Y., Yates, A.N., Delgado, A., Zhou, D., Kheyrkhah, T., Smith, J., Fiscus, J.: Mfc datasets: Large-scale benchmark datasets for media forensic challenge evaluation. In: 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), pp. 63–72. IEEE (2019)
52. Guo, C., Cao, J., Zhang, X., Shu, K., Yu, M.: Exploiting emotions for fake news detection on social media. arXiv preprint arXiv:1903.01728 (2019)
53. Gupta, A., Lamba, H., Kumaraguru, P., Joshi, A.: Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 729–736 (2013)
54. Hamdi, T., Slimi, H., Bounhas, I., Slimani, Y.: A hybrid approach for fake news detection in twitter based on user features and graph embedding. In: International Conference on Distributed Computing and Internet Technology, pp. 266–280. Springer (2020)
55. Han, Y., Karunasekera, S., Leckie, C.: Graph neural networks with continual learning for fake news detection from social media. arXiv preprint arXiv:2007.03316 (2020)
56. Heller, S., Rossetto, L., Schuldt, H.: The ps-battles dataset-an image collection for image manipulation detection. arXiv preprint arXiv:1804.04866 (2018)
57. Helmstetter, S., Paulheim, H.: Weakly supervised learning for fake news detection on twitter. In: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 274–277. IEEE (2018)
58. Hooi, B., Shah, N., Beutel, A., Günnemann, S., Akoglu, L., Kumar, M., Makhija, D., Faloutsos, C.: Birdnest: Bayesian inference for ratings-fraud detection. In: Proceedings of the 2016 SIAM International Conference on Data Mining, pp. 495–503. SIAM (2016)
59. Horne, B., Adali, S.: This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 11 (2017)
60. Hou, R., Perez-Rosas, V., Loeb, S., Mihalcea, R.: Towards automatic detection of misinformation in online medical videos. In: 2019 International Conference on Multimodal Interaction, pp. 235–243 (2019)
61. Hsu, C.C., Zhuang, Y.X., Lee, C.Y.: Deep fake image detection based on pairwise learning. Appl. Sci. **10**(1), 370 (2020)
62. Jagtap, R., Kumar, A., Goel, R., Sharma, S., Sharma, R., George, C.P.: Misinformation detection on youtube using video captions. arXiv preprint arXiv:2107.00941 (2021)
63. Jang, Y., Park, C.H., Seo, Y.S.: Fake news analysis modeling using quote retweet. Electronics **8**(12), 1377 (2019)
64. Jiang, L., Li, R., Wu, W., Qian, C., Loy, C.C.: Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2889–2898 (2020)
65. Jin, Z., Cao, J., Zhang, Y., Luo, J.: News verification by exploiting conflicting social viewpoints in microblogs. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 30 (2016)
66. Juuti, M., Sun, B., Mori, T., Asokan, N.: Stay on-topic: generating context-specific fake restaurant reviews. In: European Symposium on Research in Computer Security, pp. 132–151. Springer (2018)
67. Kaliyar, R.K., Goswami, A., Narang, P., Sinha, S.: Fndnet-a deep convolutional neural network for fake news detection. Cogn. Syst. Res. **61**, 32–44 (2020)
68. Karimi, H., Roy, P., Saba-Sadiya, S., Tang, J.: Multi-source multi-class fake news detection. In: Proceedings of the 27th International Conference on Computational Linguistics, pp. 1546–1557 (2018)
69. Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., Mora, H.: A framework for big data analytics in commercial social networks: a case study on sentiment analysis and fake review detection for marketing decision-making. Ind. Mark. Manag. **90**, 523–537 (2019)
70. Kaur, N., Jindal, N., Singh, K.: Efficient hybrid passive method for the detection and localization of copy-move and spliced images. Turk. J. Electr. Eng. Comput. Sci. **29**(2), 5 (2021)
71. Khan, J.Y., Khondaker, M., Islam, T., Iqbal, A., Afroz, S.: A benchmark study on machine learning methods for fake news detection. arXiv preprint arXiv:1905.04749 (2019)
72. Kim, J., Tabibian, B., Oh, A., Schölkopf, B., Gomez-Rodriguez, M.: Leveraging the crowd to detect and reduce the spread of fake news and misinformation. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, pp. 324–332 (2018)
73. Kinsora, A., Barron, K., Mei, Q., Vydiswaran, V.V.: Creating a labeled dataset for medical misinformation in health forums. In: 2017 IEEE International Conference on Healthcare Informatics (ICHI), pp. 456–461. IEEE (2017)
74. Kostakos, P., Nykanen, M., Martinviita, M., Pandya, A., Oussalah, M.: Meta-terrorism: identifying linguistic patterns in public discourse after an attack. In: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 1079–1083. IEEE (2018)
75. Saroop, K.P., and Augustine, R.: Fake news in social media: an analytical study on Kerala flood 2018
76. Krishnan, S., Chen, M.: Identifying tweets with fake news. In: 2018 IEEE International Conference on Information Reuse and Integration (IRI), pp. 460–464. IEEE (2018)
77. Kumar, S., Hooi, B., Makhija, D., Kumar, M., Faloutsos, C., Subrahmanian, V.: Rev2: Fraudulent user prediction in rating

platforms. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, pp. 333–341 (2018)
78. Kumar, S., Shah, N.: False information on web and social media: a survey. arXiv preprint arXiv:1804.08559 (2018)
79. Kumar, S., West, R., Leskovec, J.: Disinformation on the web: Impact, characteristics, and detection of Wikipedia hoaxes. In: Proceedings of the 25th International Conference on World Wide Web, pp. 591–602 (2016)
80. Lappas, T., Sabnis, G., Valkanas, G.: The impact of fake reviews on online visibility: a vulnerability assessment of the hotel industry. Inf. Syst. Res. **27**(4), 940–961 (2016)
81. Last, M., Markov, A., Kandel, A.: Multi-lingual detection of web terrorist content. In: Intelligence and Security Informatics, pp. 79–96. Springer (2008)
82. Lee, K., Ham, J., Yang, S.B., Koo, C.: Can you identify fake or authentic reviews? An fsQCA approach. In: Information and Communication Technologies in Tourism 2018, pp. 214–227. Springer (2018)
83. Lee, S.: Detection of political manipulation through unsupervised learning. TIIS **13**(4), 1825–1844 (2019)
84. Li, D., Guo, H., Wang, Z., Zheng, Z.: Unsupervised fake news detection based on autoencoder. IEEE Access **9**, 29356–29365 (2021)
85. Li, L., Bao, J., Zhang, T., Yang, H., Chen, D., Wen, F., Guo, B.: Face x-ray for more general face forgery detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5001–5010 (2020)
86. Li, L., Levi, O., Hosseini, P., Broniatowski, D.A.: A multi-modal method for satire detection using textual and visual cues. arXiv preprint arXiv:2010.06671 (2020)
87. Li, Y., Lyu, S.: Exposing deepfake videos by detecting face warping artifacts. arXiv preprint arXiv:1811.00656 (2018)
88. Li, Y., Yang, X., Sun, P., Qi, H., Lyu, S.C.D.: A large-scale challenging dataset for deepfake forensics. arXiv preprint arXiv:1909.12962 (2020)
89. Liao, H., Liu, Q., Shu, K., et al.: Incorporating user-comment graph for fake news detection. arXiv preprint arXiv:2011.01579 (2020)
90. Liu, Y., Wu, Y.F.B.: Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)
91. Liu, Y., Wu, Y.F.B.: Fned: a deep network for fake news early detection on social media. ACM Trans. Inf. Syst. (TOIS) **38**(3), 1–33 (2020)
92. Liu, Y., Zhu, X., Zhao, X., Cao, Y.: Adversarial learning for constrained image splicing detection and localization based on atrous convolution. IEEE Trans. Inf. Forensics Secur. **14**(10), 2551–2566 (2019)
93. Lu, Y., Zhang, L., Xiao, Y., Li, Y.: Simultaneously detecting fake reviews and review spammers using factor graph model. In: Proceedings of the 5th Annual ACM Web Science Conference, pp. 225–233 (2013)
94. Luca, M., Zervas, G.: Fake it till you make it: reputation, competition, and yelp review fraud. Manag. Sci. **62**(12), 3412–3427 (2016)
95. Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B.J., Wong, K.F., Cha, M.: Detecting rumors from microblogs with recurrent neural networks (2016)
96. Ma, T., Zhou, H., Tian, Y., Al-Nabhan, N.: A novel rumor detection algorithm based on entity recognition, sentence reconfiguration, and ordinary differential equation network. Neurocomputing **447**, 224–234 (2021)
97. Mahfoudi, G., Tajini, B., Retraint, F., Morain-Nicolier, F., Dugelay, J.L., Marc, P.: Defacto: Image and face manipulation dataset. In: 2019 27th European Signal Processing Conference (EUSIPCO), pp. 1–5. IEEE (2019)
98. Malhotra, B., Vishwakarma, D.K.: Classification of propagation path and tweets for rumor detection using graphical convolutional networks and transformer based encodings. In: 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM), pp. 183–190. IEEE (2020)
99. Marra, F., Gragnaniello, D., Cozzolino, D., Verdoliva, L.: Detection of gan-generated fake images over social networks. In: 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), pp. 384–389. IEEE (2018)
100. Meel, P., Vishwakarma, D.K.: Fake news, rumor, information pollution in social media and web: a contemporary survey of state-of-the-arts, challenges and opportunities. Expert Syst. Appl. **153**, 112986 (2020)
101. Meel, P., Vishwakarma, D.K.: Deep neural architecture for veracity analysis of multimodal online information. In: 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 7–12. IEEE (2021)
102. Meel, P., Vishwakarma, D.K.: Han, image captioning, and forensics ensemble multimodal fake news detection. Inf. Sci. **567**, 23–41 (2021)
103. Meel, P., Vishwakarma, D.K.: Machine learned classifiers for trustworthiness assessment of web information contents. In: 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), pp. 29–35. IEEE (2021)
104. Meel, P., Vishwakarma, D.K.: A temporal ensembling based semi-supervised convnet for the detection of fake news articles. Expert Syst. Appl. **177**, 115002 (2021)
105. Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., Joulin, A.: Advances in pre-training distributed word representations. arXiv preprint arXiv:1712.09405 (2017)
106. Mitra, T., Gilbert, E.: Credbank: A large-scale social media corpus with associated credibility annotations. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 9 (2015)
107. Mkono, M.: "Troll alert!": provocation and harassment in tourism and hospitality social media. Curr. Issues Tour. **21**(7), 791–804 (2018)
108. Mondal, T., Pramanik, P., Bhattacharya, I., Boral, N., Ghosh, S.: Analysis and early detection of rumors in a post disaster scenario. Inf. Syst. Front. **20**(5), 961–979 (2018)
109. Monti, F., Frasca, F., Eynard, D., Mannion, D., Bronstein, M.M.: Fake news detection on social media using geometric deep learning. arXiv preprint arXiv:1902.06673 (2019)
110. Mukherjee, A., Venkataraman, V., Liu, B., Glance, N.: What yelp fake review filter might be doing? In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 7 (2013)
111. Murayama, T., Wakamiya, S., Aramaki, E.: Mitigation of diachronic bias in fake news detection dataset. arXiv preprint arXiv:2108.12601 (2021)
112. Nataraj, L., Goebel, M., Mohammed, T.M., Chandrasekaran, S., Manjunath, B.: Holistic image manipulation detection using pixel co-occurrence matrices. arXiv preprint arXiv:2104.05693 (2021)
113. Nied, A.C., Stewart, L., Spiro, E., Starbird, K.: Alternative narratives of crisis events: Communities and social botnets engaged on social media. In: Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, pp. 263–266 (2017)
114. Nikiforos, M.N., Vergis, S., Stylidou, A., Augoustis, N., Kermanidis, K.L., Maragoudakis, M.: Fake news detection regarding the Hong Kong events from tweets. In: IFIP International Conference on Artificial Intelligence Applications and Innovations, pp. 177–186. Springer (2020)
115. Novozamsky, A., Mahdian, B., Saic, S.: Imd2020: A large-scale annotated dataset tailored for detecting manipulated images. In:

Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops, pp. 71–80 (2020)
116. Olivieri, A., Shabani, S., Sokhn, M., Cudré-Mauroux, P.: Creating task-generic features for fake news detection. In: Proceedings of the 52nd Hawaii International Conference on System Sciences (2019)
117. Paka, W.S., Bansal, R., Kaushik, A., Sengupta, S., Chakraborty, T.: Cross-SEAN: a cross-stitch semi-supervised neural attention model for covid-19 fake news detection. Appl. Soft Comput. **107**, 107393 (2021)
118. Papadopoulou, O., Zampoglou, M., Papadopoulos, S., Kompatsiaris, I.: A corpus of debunked and verified user-generated videos. Online Inf. Rev. (2019)
119. Patwari, A., Goldwasser, D., Bagchi, S.: Tathya: A multi-classifier system for detecting check-worthy statements in political debates. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 2259–2262 (2017)
120. Pérez-Rosas, V., Kleinberg, B., Lefevre, A., Mihalcea, R.: Automatic detection of fake news. arXiv preprint arXiv:1708.07104 (2017)
121. Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., Stein, B.: A stylometric inquiry into hyperpartisan and fake news. arXiv preprint arXiv:1702.05638 (2017)
122. Preston, S., Anderson, A., Robertson, D.J., Shephard, M.P., Huhe, N.: Detecting fake news on facebook: the role of emotional intelligence. PLoS ONE **16**(3), e0246757 (2021)
123. Priya, S., Sequeira, R., Chandra, J., Dandapat, S.K.: Where should one get news updates: Twitter or reddit. Online Soc. Netw. Media **9**, 17–29 (2019)
124. Proferes, N., Jones, N., Gilbert, S., Fiesler, C., Zimmer, M.: Studying reddit: a systematic overview of disciplines, approaches, methods, and ethics. Soc. Media Soc. **7**(2), 20563051211019004 (2021)
125. Qian, S., Hu, J., Fang, Q., Xu, C.: Knowledge-aware multi-modal adaptive graph convolutional networks for fake news detection. ACM Trans. Multimed. Comput. Commun. Appl. (TOMM) **17**(3), 1–23 (2021)
126. Rajdev, M., Lee, K.: Fake and spam messages: Detecting misinformation during natural disasters on social media. In: 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), vol. 1, pp. 17–20. IEEE (2015)
127. Rastogi, S., Bansal, D.: Visualization of twitter sentiments on Kashmir territorial conflict. Cybern. Syst. **52**(8), 642–669 (2021)
128. Rastogi, S., Bansal, D.: Disinformation detection on social media: an integrated approach. Multimed. Tools Appl. **81**, 1–33 (2022)
129. Reis, J., Melo, P.d.F., Garimella, K., Benevenuto, F.: Can whatsapp benefit from debunked fact-checked stories to reduce misinformation? arXiv preprint arXiv:2006.02471 (2020)
130. Reis, J.C., Correia, A., Murai, F., Veloso, A., Benevenuto, F.: Supervised learning for fake news detection. IEEE Intell. Syst. **34**(2), 76–81 (2019)
131. Ribeiro, M.H., Calais, P.H., Almeida, V.A., Meira Jr, W.: " everything i disagree with is# fakenews": correlating political polarization and spread of misinformation. arXiv preprint arXiv:1706.05924 (2017)
132. Riedel, B., Augenstein, I., Spithourakis, G.P., Riedel, S.: A simple but tough-to-beat baseline for the fake news challenge stance detection task. arXiv preprint arXiv:1707.03264 (2017)
133. Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., Nießner, M.: Faceforensics++: Learning to detect manipulated facial images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1–11 (2019)
134. Rubin, V.L., Chen, Y., Conroy, N.K.: Deception detection for news: three types of fakes. Proc. Assoc. Inf. Sci. Technol. **52**(1), 1–4 (2015)
135. Rubin, V.L., Conroy, N., Chen, Y., Cornwell, S.: Fake news or truth? using satirical cues to detect potentially misleading news. In: Proceedings of the Second Workshop on Computational Approaches to Deception Detection, pp. 7–17 (2016)
136. Ruchansky, N., Seo, S., Liu, Y.: 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 (2017)
137. Sabir, E., Cheng, J., Jaiswal, A., AbdAlmageed, W., Masi, I., Natarajan, P.: Recurrent convolutional strategies for face manipulation detection in videos. Interfaces (GUI) **3**(1), 80–87 (2019)
138. Saez-Trumper, D.: Online disinformation and the role of wikipedia. arXiv preprint arXiv:1910.12596 (2019)
139. Saffi, H., Do, T.P., Hansen, J.M., Dodick, D.W., Ashina, M.: The migraine landscape on youtube: a review of youtube as a source of information on migraine. Cephalalgia **40**(12), 1363–1369 (2020)
140. Sahoo, S.R., Gupta, B.B.: Multiple features based approach for automatic fake news detection on social networks using deep learning. Appl. Soft Comput. **100**, 106983 (2021)
141. Saikh, T., De, A., Ekbal, A., Bhattacharyya, P.: A deep learning approach for automatic detection of fake news. arXiv preprint arXiv:2005.04938 (2020)
142. Sánchez-Rebollo, C., Puente, C., Palacios, R., Piriz, C., Fuentes, J.P., Jarauta, J.: Detection of jihadism in social networks using big data techniques supported by graphs and fuzzy clustering. Complexity **2019** (2019)
143. Schuckert, M., Liu, X., Law, R.: Insights into suspicious online ratings: direct evidence from TripAdvisor. Asia Pac. J. Tour. Res. **21**(3), 259–272 (2016)
144. Shahi, G.K.: Amused: an annotation framework of multi-modal social media data. arXiv preprint arXiv:2010.00502 (2020)
145. Shahi, G.K., Nandini, D.: Fakecovid–a multilingual cross-domain fact check news dataset for covid-19. arXiv preprint arXiv:2006.11343 (2020)
146. Shan, G., Zhou, L., Zhang, D.: From conflicts and confusion to doubts: examining review inconsistency for fake review detection. Decis. Support Syst. **144**, 113513 (2021)
147. Shang, L., Zhang, D.Y., Wang, M., Lai, S., Wang, D.: Towards reliable online clickbait video detection: a content-agnostic approach. Knowl.-Based Syst. **182**, 104851 (2019)
148. Shao, C., Ciampaglia, G.L., Varol, O., Flammini, A., Menczer, F.: The spread of fake news by social bots. arXiv preprint arXiv:1707.07592 **96**, 104 (2017)
149. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake news detection on social media: a data mining perspective. ACM SIGKDD Explor. Newsl. **19**(1), 22–36 (2017)
150. Shu, K., Wang, S., Liu, H.: Exploiting tri-relationship for fake news detection. arXiv preprint arXiv:1712.07709 **8** (2017)
151. Sicilia, R., Giudice, S.L., Pei, Y., Pechenizkiy, M., Soda, P.: Twitter rumour detection in the health domain. Expert Syst. Appl. **110**, 33–40 (2018)
152. Siering, M., Koch, J.A., Deokar, A.V.: Detecting fraudulent behavior on crowdfunding platforms: the role of linguistic and content-based cues in static and dynamic contexts. J. Manag. Inf. Syst. **33**(2), 421–455 (2016)
153. Silva, A., Luo, L., Karunasekera, S., Leckie, C.: Embracing domain differences in fake news: Cross-domain fake news detection using multi-modal data. arXiv preprint arXiv:2102.06314 (2021)
154. Silverman, C.: This analysis shows how viral fake election news stories outperformed real news on facebook. BuzzFeed news **16** (2016)
155. Singh, S.: Everything in moderation: An analysis of how internet platforms are using artificial intelligence to moderate user-generated content. New America **22** (2019)
156. Singhal, S., Shah, R.R., Chakraborty, T., Kumaraguru, P., Satoh, S.: Spotfake: A multi-modal framework for fake news detection.

In: 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), pp. 39–47. IEEE (2019)

157. Singhania, S., Fernandez, N., Rao, S.: 3han: a deep neural network for fake news detection. In: International Conference on Neural Information Processing, pp. 572–581. Springer (2017)
158. Situngkir, H.: Spread of hoax in social media (2011)
159. Solorio, T., Hasan, R., Mizan, M.: A case study of sockpuppet detection in wikipedia. In: Proceedings of the Workshop on Language Analysis in Social Media, pp. 59–68 (2013)
160. Tacchini, E., Ballarin, G., Della Vedova, M.L., Moret, S., de Alfaro, L.: Some like it hoax: Automated fake news detection in social networks. arXiv preprint arXiv:1704.07506 (2017)
161. Thu, P.P., New, N.: Implementation of emotional features on satire detection. In: 2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), pp. 149–154. IEEE (2017)
162. Van Bruwaene, D., Huang, Q., Inkpen, D.: A multi-platform dataset for detecting cyberbullying in social media. Lang. Resour. Eval. **54**(4), 851–874 (2020)
163. Varshney, D., Vishwakarma, D.K.: Analysing and identifying crucial evidences for the prediction of false information proliferated during covid-19 outbreak: a case study. In: 2021 8th International Conference on Smart Computing and Communications (ICSCC), pp. 47–51. IEEE (2021)
164. Varshney, D., Vishwakarma, D.K.: Artimarker: A novel artificially inflated video marking and characterization method on youtube. In: 2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP), pp. 244–249. IEEE (2021)
165. Varshney, D., Vishwakarma, D.K.: An automated multi-web platform voting framework to predict misleading information proliferated during covid-19 outbreak using ensemble method. arXiv preprint arXiv:2109.13063 (2021)
166. Varshney, D., Vishwakarma, D.K.: Hoax news-inspector: a real-time prediction of fake news using content resemblance over web search results for authenticating the credibility of news articles. J. Ambient. Intell. Humaniz. Comput. **12**(9), 8961–8974 (2021)
167. Varshney, D., Vishwakarma, D.K.: A review on rumour prediction and veracity assessment in online social network. Expert Syst. Appl. **168**, 114208 (2021)
168. Varshney, D., Vishwakarma, D.K.: A unified approach for detection of clickbait videos on youtube using cognitive evidences. Appl. Intell. **51**(7), 4214–4235 (2021)
169. Vincent, N.S.Y., Pal, A., Chua, A.Y.: Studying healthcare personnel's intention to click clickbaits. In: Proceedings of the International MultiConference of Engineers and Computer Scientists, vol. 1 (2018)
170. Vishwakarma, D.K., Jain, C.: Recent state-of-the-art of fake news detection: a review. In: 2020 International Conference for Emerging Technology (INCET), pp. 1–6. IEEE (2020)
171. Vishwakarma, D.K., Varshney, D., Yadav, A.: Detection and veracity analysis of fake news via scrapping and authenticating the web search. Cogn. Syst. Res. **58**, 217–229 (2019)
172. Vizoso, Á., Vaz-Álvarez, M., López-García, X.: Fighting deepfakes: media and internet giants' converging and diverging strategies against Hi-tech misinformation. Media Commun. **9**(1), 291–300 (2021)
173. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science **359**(6380), 1146–1151 (2018)
174. Vuković, M., Pripužić, K., Belani, H.: An intelligent automatic hoax detection system. In: International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, pp. 318–325. Springer (2009)
175. Wang, B., Zhuang, J.: Rumor response, debunking response, and decision makings of misinformed twitter users during disasters. Nat. Hazards **93**(3), 1145–1162 (2018)
176. Wang, L., Shen, X., de Melo, G., Weikum, G.: Cross-domain learning forclassifying propaganda in online contents. arXiv preprint arXiv:2011.06844 (2020)
177. Wang, W.Y.: " liar, liar pants on fire": A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648 (2017)
178. Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Su, L., Gao, J.: Eann: Event adversarial neural networks for multi-modal fake news detection. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 849–857 (2018)
179. Wardle, C., Derakhshan, H.: Information disorder: toward an interdisciplinary framework for research and policy making. Counc. Eur. Rep. **27**, 1–107 (2017)
180. Wu, K., Yang, S., Zhu, K.Q.: False rumors detection on sina weibo by propagation structures. In: 2015 IEEE 31st International Conference on Data Engineering, pp. 651–662. IEEE (2015)
181. Xing, J., Wang, S., Zhang, X., Ding, Y.: HMBI: a new hybrid deep model based on behavior information for fake news detection. Wirel. Commun. Mob. Comput. (2021). https://doi.org/10.1155/2021/9076211
182. Yang, X., Li, Y., Lyu, S.: Exposing deep fakes using inconsistent head poses. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8261–8265. IEEE (2019)
183. Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., Yu, P.S.: Ti-cnn: Convolutional neural networks for fake news detection. arXiv preprint arXiv:1806.00749 (2018)
184. Yoo, K.H., Gretzel, U.: Comparison of deceptive and truthful travel reviews. In: ENTER, pp. 37–47 (2009)
185. Zaman, B., Justitia, A., Sani, K.N., Purwanti, E.: An Indonesian hoax news detection system using reader feedback and naïve bayes algorithm. Cybern. Inf. Technol. **20**(1), 82–94 (2020)
186. Zannettou, S., Caulfield, T., De Cristofaro, E., Kourtelris, N., Leontiadis, I., Sirivianos, M., Stringhini, G., Blackburn, J.: The web centipede: understanding how web communities influence each other through the lens of mainstream and alternative news sources. In: Proceedings of the 2017 Internet Measurement Conference, pp. 405–417 (2017)
187. Zannettou, S., Sirivianos, M., Blackburn, J., Kourtellis, N.: The web of false information: rumors, fake news, hoaxes, clickbait, and various other shenanigans. J. Data Inf. Qual. (JDIQ) **11**(3), 1–37 (2019)
188. Zhang, X., Ghorbani, A.A.: An overview of online fake news: characterization, detection, and discussion. Inf. Process. Manag. **57**(2), 102025 (2020)
189. Zhao, Z., Zhao, J., Sano, Y., Levy, O., Takayasu, H., Takayasu, M., Li, D., Wu, J., Havlin, S.: Fake news propagates differently from real news even at early stages of spreading. EPJ Data Sci. **9**(1), 7 (2020)
190. Zheng, H.T., Chen, J.Y., Yao, X., Sangaiah, A.K., Jiang, Y., Zhao, C.Z.: Clickbait convolutional neural network. Symmetry **10**(5), 138 (2018)
191. Zhou, P., Han, X., Morariu, V.I., Davis, L.S.: Two-stream neural networks for tampered face detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1831–1839. IEEE (2017)
192. Zhou, X., Cao, J., Jin, Z., Xie, F., Su, Y., Chu, D., Cao, X., Zhang, J.: Real-time news cer tification system on sina weibo. In: Proceedings of the 24th International Conference on World Wide Web, pp. 983–988 (2015)
193. Zhou, X., Jain, A., Phoha, V.V., Zafarani, R.: Fake news early detection: a theory-driven model. Digit. Threats: Res. Pract. **1**(2), 1–25 (2020)
194. Zhou, X., Wu, J., Zafarani, R.: Safe: Similarity-aware multi-modal fake news detection. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 354–367. Springer (2020)

195. Zhou, X., Zafarani, R.: Network-based fake news detection: a pattern-driven approach. ACM SIGKDD Explor. Newsl. **21**(2), 48–60 (2019)
196. Zhou, X., Zafarani, R.: A survey of fake news: fundamental theories, detection methods, and opportunities. ACM Comput. Surv. (CSUR) **53**(5), 1–40 (2020)
197. Zubiaga, A., Liakata, M., Procter, R., G, Wong Sak Hoi., Tolmie, P.: Analysing how people orient to and spread rumours in social media by looking at conversational threads. PLoS ONE **11**(3), e0150989 (2016)