



A comprehensive review on automatic detection of fake news on social media

Manish Kumar Singh¹ · Jawed Ahmed¹ · Mohammad Afshar Alam¹ · Kamlesh Kumar Raghuvanshi² · Sachin Kumar³

Received: 8 February 2023 / Revised: 27 June 2023 / Accepted: 1 October 2023 /

Published online: 26 October 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Social media sites are now quite popular among internet users for sharing news and opinions. This has become possible due to the inexpensive Internet, easy availability of digital devices, and the no-cost policy to create a user account on social media platforms. People are fascinated by social media sites because they can easily connect with others to share their interests, news, and opinions. According to studies, someone who lacks credibility is more likely to spread false information in order to achieve goals of any kind, be it influencing political opinions, earning attention, or making money. The automatic detection of social media related fake news has thus emerged as a highly anticipated research area in recent years. This paper offers a comprehensive review of the automatic detection of fake news on social media platforms. It contains details of the key models or techniques related to machine/deep learning proposed (or developed) during the period of the year 2011 to the year 2022 along with the performance metrics of each model or technique. The paper discusses (a). the key challenges faced during the development of an effective and efficient fake news detection system, (b). some popular datasets for carrying out fake news detection related research, and (c). the major research gaps, and future research directions in the area of automatic fake news detection on social media.

Keywords Social media · Fake news · Detection methods · Machine learning · Datasets

1 Introduction

Nowadays, social media sites are used as a form of democratic media by internet users to often exchange opinions and news [1]. This is particularly true in developing nations where sixty-two percent of users get their news through social media [2]. This has become possible due to improved technology, easy availability of digital devices, inexpensive Internet, and the no-cost policy to create a user account on social media platforms [3]. Consequently, conventional news media like newspapers, magazines, television, radio, etc. are switching

✉ Manish Kumar Singh
manish.kumar2191989@gmail.com; manishksingh_sch@jamiyahamdard.ac.in

Extended author information available on the last page of the article

to digital platforms like blogs, news sites, and social media platforms (Twitter, Facebook, etc.) to reach a large audience in different ways [4]. Social media platforms are better able to connect with huge audiences as compared to conventional media and other digital media [5]. Furthermore, social media may be a tremendous tool for individuals and businesses to interact and provide details in less time than ever before if used properly. However, the information that consumers acquire via social media sites is obviously not always accurate and might not reach them right away. What's worse is that these sites are being utilized to disseminate misleading information, which is appropriately referred to as "fake news" [6], for the purpose of influencing others' thoughts solely for one's own benefit [4].

The proliferation of news portals and social media sites propagating false news is driven by two key factors [2]: *a). intentions to make money*, that entail generating a sizable quantity of income from viral news stories; and *b). intentions to build particular ideology*, that focus on influencing people's views on particular issues. Furthermore, the proliferation of bots and trolls, which are online malevolent agents, is one of the major factors in the spread of false information [7] in the present era.

Fake news has existed for a while [8]. Long before the Internet was invented, there were fake news and hoaxes [9]. It is crucial right now to inquire as to why it has gained global notice and why there is an increase in public awareness about fake news. The primary cause of this is that online fake news production and publication are cheaper and faster as compared to conventional news media like newspapers and television. Social media hoaxes, rightly known as fake news, have raised considerable concerns across the globe. It is clear that a layperson, due to a lack of expertise, may struggle to determine the validity of fake news. As a result, in recent years, academics, researchers, and the industry have been inclined to the development of effective and reliable techniques for spotting false news on social media. In other words, limiting the required work and time by an individual to recognize false news and restrict its spread are the prime goals of fake news detection research. In order to make early as well as automatic spotting of social media related fake news, several attempts have been made so far, particularly in the area of machine/deep learning [10].

1.1 Research contribution

The following is a list of this paper's main contributions:

1. To review and provide a summary of 60 key published papers related to automatically detecting fake news on social media platforms;
2. To provide the details of key models or techniques related to machine/deep learning that was proposed (or developed) during the period of the year 2011 to the year 2022 and present their performance metric;
3. To discuss key challenges faced during the development of a reliable approach for spotting false news and provide details of some popular datasets for carrying out research related to spotting fake news; and
4. To present gaps in the literature and potential future directions in the field of automatic detection of social media related fake news.

1.2 Paper organization

There are five sections in the paper. The theoretical background for detecting news on social media platforms is presented in Section 2; Section 3 outlines the research methodology

used to conduct the current review; Section 4 provides the summary of 60 key published papers related to the automatic detection of false news on social media that were published between the time frame of the year 2011 to 2022; Section 5 covers results and discussion; and Section 6 provides the conclusion.

2 Theoretical background

The current section puts light on the fundamentals associated with the detection of fake news on social media including the definition of fake news and its detection techniques.

2.1 Definition of fake news

There is a long history to the term "fake news." But it gained popularity and developed a bad reputation throughout the 2016 election campaign for the US Presidential Post [11]. Since then, fake news has become a highly anticipated area of research, and it is widely understood to refer to the tainting of information with rumors, hoaxes, propaganda, false news, and other types of incorrect information [12]. Even in journalism, there is presently no agreement on how to define or classify fake news, despite the fact that these phrases indicate misleading information. It is therefore necessary to have a clear and correct definition of fake news in order to examine it and evaluate relevant studies related to it. In order to get a solid knowledge of the concept of fake news, it is crucial to first understand a few related phrases that connect to fake news prevailing on social media:

- **Hoax:** It implies a fabrication designed to appear to be the truth [13].
- **Junk News:** In general, it implies the whole news material that relates to a publication rather than simply one piece, and it gathers a variety of facts [14].
- **False News:** Untrue news reports that are meant for misleading the masses intentionally or unintentionally are known as false news. They have a similar format to that of conventional news articles but differ in their organization and goal [12, 15].
- **Click-bait:** It is regarded as being low-quality journalism done to draw viewers and bring in money from advertising [16, 17].
- **Rumor:** According to certain academicians, rumor falls within the genre of propaganda [18]. It is derived from a Latin word that implies "noise". It is regarded as an unverified statement propagated from one person to another [19].
- **Satire:** It implies a kind of news that is produced with the intention of criticizing or entertaining masses that may seem to be actual news but can be harmful if shared inappropriately [8, 20].
- **Propaganda:** It suggests the use of deception, selective omission, or the presentation of biased information in news stories to affect the target audience's feelings, thoughts, and actions in order to advance certain ideologies, causes, or beliefs [8].

All the above terms associated with fake news can be categorized into three types of information [21] that are compared in Table 1:

- **Malinformation:** It is described as a factual assertion that is used to harm a person, a group, or a country [22].
- **Disinformation:** It consists of the purposeful, deliberate, or intentional spread of inaccurate or misleading information with the objective to confuse, deceive, or mislead the public [23, 24].

Table 1 Associated terms of fake news

Associated term	Malinformation	Disinformation	Misinformation
Hoax	No	Yes	Yes
Junk News	Yes	Yes	No
False News	Yes	Yes	Yes
Click-bait	No	Yes	No
Rumor	No	Yes	Yes
Satire	Yes	No	No
Propaganda	Yes	Yes	No

- **Misinformation:** It can be described as the spread of misleading information without an agenda by those who are either ignorant of the truth or have cognitive impairments [25, 26].

The increasing digitalization of news has changed the perception towards news due to which the traditional way of defining news won't work. Zhou, Zafarani [27], thus, defined fake news broadly as:

Definition 2.1 Fake news is false news

According to the above description, fake news generally contains declarations, assertions, posts, articles, speeches, and other material regarding a certain topic. Furthermore, such news can be produced by both professional and non-professional journalists.

Shu et al. [6] describe fake news in more detail while making sure it satisfies the general standards for being false news and covers the public's perception of fake news:

Definition 2.2 A news piece that has been purposefully written with the objective to mislead and that can be verified as false is called fake news.

2.2 The impact of fake news on the real world

The rapid increase in interest in fake news can be explained by taking into account a lot of significant incidents that have recently occurred throughout the world. According to Vosoughi et al. [15], the majority of cases involving fake news are actually motivated by politics. Following the 2016 US presidential elections, the phrase "fake news" has become a buzzword [28]. According to some, it's possible that Donald Trump didn't win the election had it not been for the consequences of misleading information (besides the alleged Russian troll influence) [2]. Furthermore, according to a study, fake news influenced the 2016 UK Brexit referendum [29]. Some other noteworthy fake news events on a global scale include the stock market crash caused by a false tweet about the death of President Obama in 2013 [30], the false reports on Pizzagate that caused an attack on a restaurant with gunfire [6], and the general doubt related to vaccines prevailed during the epidemics of Ebola and Zika virus [31]. The publication of a fake news item that hackers had plotted in a news agency in the Middle East nation of Qatar was one of the factors that caused the 2017 Gulf diplomatic impasse, which lasted three months [32]. Ideological changes are commonly brought on by the spread of false information in the Middle East [33]. Nearly two weeks following Russia's invasion of Ukraine in 2022, numerous erroneous reports claimed that the conflict was a fraud, a media invention, or that the West had exaggerated its severity [34].

In another instance, the Indian PM's declaration of a Janata curfew in March 2020 led to a national lockdown for stopping the spread of COVID-19. This became the subject of research on misinformation in India, as shown in Fig. 1, which revealed a spike in the number of untrue stories [35]. Twenty-two YouTube news channels were reportedly blocked by the Indian government for allegedly providing false information in an effort to deceive viewers [36]. Since there are far too many real-time instances of fake news, they cannot all be included here.

2.3 Fake news detection techniques

In order to identify whether a piece of news is fake or not, it is important to first develop an understanding of essential concepts related to fake news detection including its mathematical modeling, framework, model development, and prospective dataset for the model's training and testing. The mathematical model for the identification of fake news is a part of the machine learning binary classification problem. In a data mining framework for spotting fake news, features are extracted, then a model is built. During the feature extraction stage, a formalized mathematical framework representing news content and associated auxiliary data are created. News content, social context, and domain-specific feature extraction are the three most widely used techniques for identifying social media related fake news. Figure 2 provides the broad categorization of feature extraction types.

2.2.1. News-Content-based feature extraction Features of news content explain the meta-data for a particular piece of news. It is possible to build various feature representations depending on the raw content properties (like the author, title, body, and any embedded images or videos in a news story) to produce discriminatory elements of false news using news content. The news stories that are frequently utilized for extracting features can be either of *linguistic* or *visual* form. By collecting various writing styles and spectacular headlines, linguistic-based elements can be used to identify false news [17]. The text content

Increase In The Number of (Debunked) Misinformation
January to April 2020

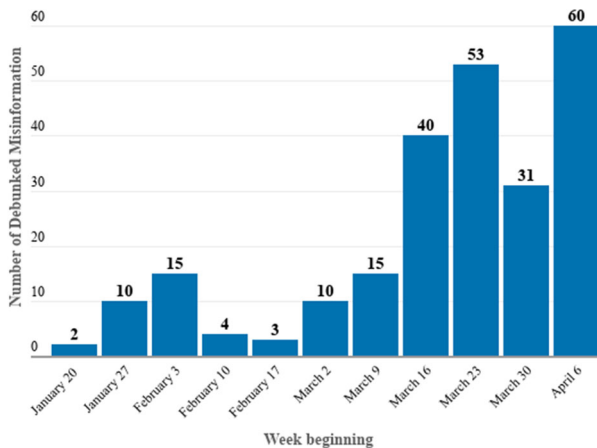


Fig. 1 An analysis of rumors spread during the Janta Curfew in India (Courtesy: University of Michigan study on false information, released in April 2020)

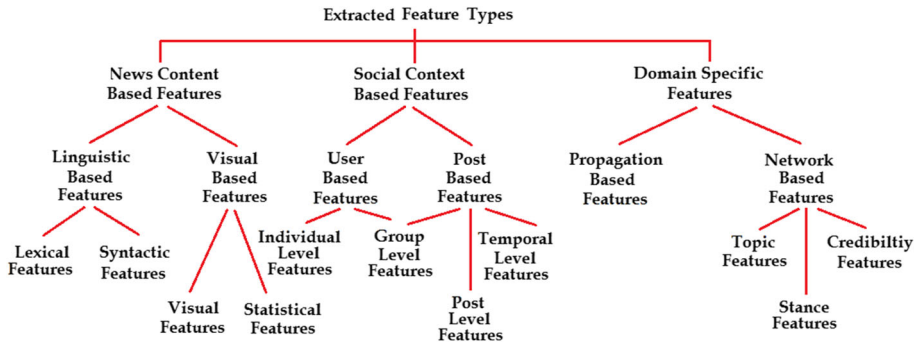


Fig. 2 The broad categorization of feature extraction types

can be used to extract linguistic-based properties in the following ways: (a). *lexical features* (includes characteristics at the levels of character and word, including the number of characters per word, unique words, the overall number of words, and the number of times long words appear), and (b). *syntactic features* (composed of sentence-level components, including widely used BoW (bag-of-words), "n-grams", POS tagging, and many more). Contrarily, fake news can be identified through visual aspects by locating its various features utilizing visual resources like photographs and videos. For checking the news veracity, several *visual features* and *statistical features* can be retrieved [37].

2.2.2 Social-Context-based feature extraction Social media interactions serve as a source of auxiliary data for determining the veracity of news articles and can serve as an easy way to illustrate the process of news diffusion over time. The two crucial components of the social media context are *users* and *posts* whose retrieval is helpful in the identification of false news. The *user-based feature extraction* involves obtaining user-profiles and characteristics that may be found either at *individual level* (by determining the credibility of the person by looking at their amount of posts, followers, etc. [10]), or at *group level* (by determining "average followers" and "percentage of verified followers" [38, 39] which serve to list the common characteristics of user groups who are interested in specific news). Contrarily, the *post-based feature extraction* involves locating user responses to social media posts in order to identify any fake news that may be done at (a). *post level* (using "feature values" for each post that can be retrieved using a number of embedding strategies in addition to the linguistic-based approach [40]), (b). *group level* (by determining the average credibility score for specific news articles utilizing the crowd's intelligence, which is the "feature value" of all noteworthy posts as a whole [41]), and (c). *temporal level* (by determining "feature values" based on the temporal variations of those features [39]).

2.2.3. Domain-Specific-based feature extraction These characteristics, which correspond to the news of specific areas, could consist of external links, paraphrased text, the count of graphs, and their mean size. They can be obtained using propagation-based or network-based methods. The *propagation-based feature extraction* determines the news veracity by using the idea of how false news spreads [42]. Contrarily, the *network based feature extraction* includes identifying and extracting features from various networks (that develop on social media sites based on common interest) including diffusion network [38], co-occurrence network [40], stance network [41] and many more.

Following the extraction of features for fake news detection, a process known as "model development" is carried out to create machine learning models that can better distinguish

between fake and real news using the data from the collected features. According to the kind of feature retrieved from the news posted on social media sites, there are two kinds of models for spotting false news, namely (i). *based on news content* (it is developed using features derived from news content), and (ii). *based on social context* (it is created utilizing collected information from social context). There are two methods for building models based on news content: based on knowledge, and based on style. The *knowledge-based technique* tries to verify claims made in a certain news report in a specific context by fact-checking proposed statements in the article utilizing external sources [43]. Building a model to spot fake news based on knowledge may require three different fact-checking techniques [27]: (a). *expert-oriented based* (it uses human subject-matter specialists from specific fields to consult important records in order to confirm the veracity of a claim made in the news), (b). *crowd-sourced based* (it uses a sizable population of persons who can annotate supplied news, and then do a general assessment of each annotation to determine whether the claim made in the news is true), and (c). *automatic* (it builds an autonomous, scalable system to check the veracity of the assertion in the supplied news using information retrieval, machine learning, graph theory, and natural language processing techniques [44]). The above-mentioned fact-checking techniques are illustrated in detail in Table 2. The *style-based approach* for building a model related to detecting fake news focus on tracking down the writers of manipulative news articles. Building a style-based model for spotting false news has so far relied on two different strategies, namely (a). *deception-based* (it works by utilizing deep network models, like those used by CNN, or rhetorical structure theory to identify false or misleading statements or claims in news reports [45]), and (b). *objectivity-based* (it recognizes style cues, such as click-bait and hyper-partisanship, that could be a sign of a decrease in the quality of news reporting in an effort to deceive the masses).

For building models based on social context to spot false news, there can be two methods: based on stance, and based on propagation. The *stance-based technique* uses direct feedback, such as "thumbs up", "thumbs down" or other reactions on social media sites in order to establish the user's stance on a specific person, thing, or idea - whether they are in favor of it, against it, or neutral [47]. Contrarily, the *propagation-based technique* employs the theory that the credibility of major posts on a social media site is directly tied to the credibility of the relevant news with the purpose of confirming the news veracity.

3 Research methodology

The research methodology followed in this paper was developed after referring to the PRISMA-defined research procedure [48] in order to review important papers related to the

Table 2 Illustrations of fact-checking techniques

Fact-check sites	Technique employed	Domain	Kind of work
Snopes ¹	Expert-oriented	Politics, social issues	Fact-checking website
PolitiFact ²	Expert-oriented	US politics	Fact-checking website
Fiskkit ³	Crowd-sourced	Politics, society, religion etc.	Discussion forum
FAKTA [46]	Automatic	MBFC website's news media	Fact-checking system

¹ <https://www.snopes.com>

² <https://www.politifact.com>

³ <https://fiskkit.com>

automatic detection of social media related false news. Five steps are taken to complete the review, and they are all clearly shown in Fig. 3:

Step-1: IEC-compliant organisation of published documents (Inclusion/Exclusion Criteria)

- Criteria 1: All cited published papers must be in English;
- Criteria 2: All cited published papers must be original works;
- Criteria 3: All cited published papers place a strong emphasis on research into the automatic detection of social media based false news; and
- Criteria 4: All cited published papers must be from a particular time frame. The current review is based on papers published from the year 2011 to 2022.

Step-2: Choosing a source to access published papers

- The initial finding for articles was done as per the keyword search of "automatic detection of social media related fake news" on Google Scholar and electronic databases like IEEEExplore, Springerlink, ScienceDirect, and ACM.
- There were ten keywords, ranging from short to long phrases, used during the search for suitable published papers in the above-discussed databases. These keywords include "automatic detection of social media related fake news", "AI techniques for detecting fake news", "ML methods for detection of fake news", "Deepfakes", "fake news datasets", "content-based detection of fake news", "context-based detection of fake news", "network-based spotting of false news", "review paper on fake news", and "review paper on fake news detection". Around 561 articles were obtained.
- The articles were then filtered on the basis of title as well as publisher, including IEEE, Springer, ACM, Elsevier, AAI, and ACL. Around 120 published papers were selected from journals with good impact factors and highly reputed conference proceedings.
- A distribution chart of the percentage of papers published by reputed publishers is presented in Fig. 4. The maximum number of articles have been published by ACM (31.0%) while ACL (6.0%) published the least number of articles.

Step-3: Selection of papers for review

The abstracts of each selected article were read to determine their relevance. On the basis of relevance, 60 articles in total were chosen for the review in Section 4.

Step-4: Data collection

All the collected data were analyzed and used in Section 5 for results and discussion.

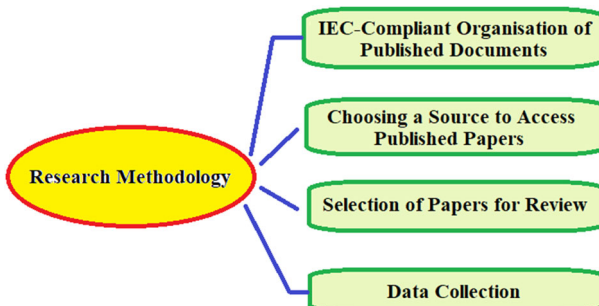


Fig. 3 Research methodology followed during review of papers

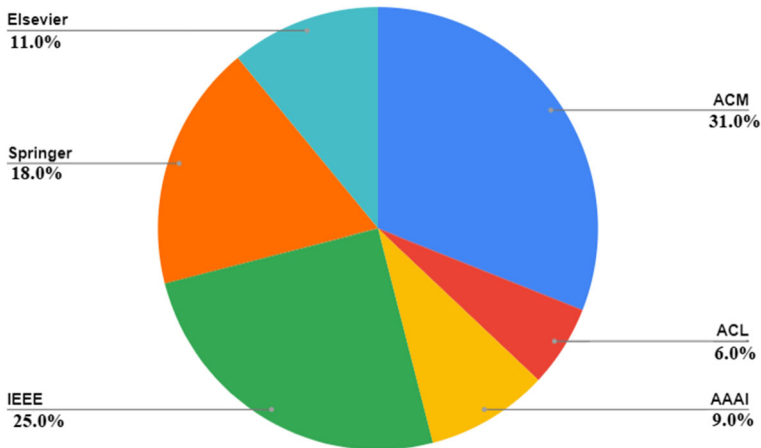


Fig. 4 Percentage of papers published by reputed publishers

4 Review of selected published papers

This section provides a comprehensive review of a total of 60 highly cited published papers in the field of automatic spotting of social media related fake news. This section first reviews all the related survey papers, followed by a review of all the research papers in order to develop a good understanding of the given research domain.

4.1 Review of survey papers

Vlachos and Riedel [43] surveyed prominent websites of fact-checking and selected two websites out of them (Channel 4 and PolitiFact) to construct the first public dataset for carrying out the fact-checking related task. They explained the procedure of fact-checking and used baseline approaches for the given task with a major limitation that decisions obtained through baseline approaches weren't easily interpreted.

Özgöbek and Gulla [49] reviewed several articles to identify key machine learning as well as linguistic methods for detecting fake news on an automated basis and they grouped them into three broad categories - content-related methods, user-related methods, and network-related methods.

Shu et al. [6] provided a survey paper on spotting social media related false news from the perspective of data mining. They not only narrowed down the idea of fake news with respect to data mining research but also provided key approaches, different models, prominent datasets as well as major evaluation metrics required for the automatic spotting of false news.

Meinert et al. [50] surveyed eight journal articles and compared them on the basis of the techniques for identifying false news on social media including social context features, credibility features, semantic features, linguistic features, network features, statistical image features, and visual features. They discussed the increasing role of social media in propagating false news.

Kumar and Shah [51] provided a thorough review of research related to fake news on social media and the web, and categorized the review into three categories - (a). on the basis of platforms of study (Twitter, Facebook, Review Platforms, Sina Weibo, Multi-Platform,

and Other); (b). on the basis of characteristics of fake news (text, user, graph, rating score, time, and propagation); and (c). on the basis of the type of algorithms used for detecting fake news (based on features, models, and graphs).

Pariikh et al. [52] provided a survey of the most recent techniques for the automatic identification of false news and described popular datasets for studying fake news events. The paper presented four key open challenges for research related to the detection of fake news.

Sharma et al. [53] provided a comprehensive review on identification as well as mitigation techniques for dealing social media related false news. The paper not only discussed the key players involved during the fake news propagation but also highlighted the serious challenges posed by the increased instances of fake news across the globe and thus presented the goals for fake news research. All the identification techniques of fake news were discussed in detail and the mitigation techniques were deeply analyzed. Finally, the distinguishing qualities of the accessible datasets for fake news research were summarized along with future work.

Pierri and Ceri [12] presented a comprehensive data-driven survey on social media related fake news. Their work focused not only on the feature classification used by each study for the identification of false information but also on the datasets for guiding classification techniques.

Da Silva et al. [54] reviewed a large number of articles to identify key machine-learning techniques for spotting fake news. Further, the paper found that the combination of classical techniques along with a neural network can detect fake news more effectively as compared to a single classical machine learning method. The paper also traced the requirement of a domain ontology in order to unify myriad definitions and terminology related to the domain of fake news so as to prevent any kind of misleading opinions and inferences.

Zhang and Ghorbani [55] presented a comprehensive review to determine the ill effects of online fake news, and identify the sophisticated detection techniques. The paper argued that most of the detection techniques are built on features extracted from the users, content, and context associated with news. The paper also covered existing datasets for detecting fake news and proposes key research directions for the future analysis of online fake news.

Mridha et al. [56] provided a comprehensive review on spotting fake news using deep learning techniques. The paper highlighted the myriad consequences of fake news, briefly tabulated major datasets for experimental setups, and presented significant approaches based on deep learning to spot fake news CNN (Convolutional Neural Network), RNN (Recurrent Neural Network), GNN (Graph Neural Network), GAN (Generative Adversarial Network), Attention Mechanism, BERT (Bidirectional Encoder Representations for Transformers), and ensemble techniques. The paper discussed the key scope of future research in the given domain.

Bani-Hani et al. [57] developed an OWL (Web Ontology Language) based Fandet semantic model to present a thorough analysis of the context-based techniques for spotting social media related fake news.

4.2 Review of research papers

This subsection provides the summary of key research articles published during the period between the years 2011 and 2022. Each summary includes the details of the model or technique developed based on machine/deep learning, the data inputs used during the experimental setup, and the performance metrics of each model or technique.

Castillo et al. [10] created a model based on a decision-tree algorithm with a broad feature set to predict the credibility of news. They used Twitter events as input data for their experiment and an accuracy of 86 percent was attained using the supervised classifier. The strength of the model is that it utilizes a broad feature set and achieves a high accuracy of 86%. It has the weakness that it is limited to Twitter events, and there is no comparison with other models.

Rubin et al. [58] categorized news as true or fraudulent using an SVM classifier using a combination of Rhetorical Structure Theory and Vector Space Modeling techniques. They utilized news transcripts with associated RST analyses for input data and the SVM classifier achieved 63% accuracy. They found their predictive model not better than the chance of 56% accuracy. However, the model was shown to outperform the fake detection abilities of humans by 54% accuracy. The model aimed to discover differences in systematic language and to report the key method that could be employed for checking news veracity by the system.

Ferreira et al. [59] utilized multiclass logistic regression to create a model for the classification of stance and provided rumored claims as well as news articles from the Emergent dataset to the system as input data. The accuracy of the model was 73 percent, which was 26 percent greater than that of the Excitement Open Platform [60].

Jin et al. [41] employed conflicting social opinions in a credibility propagation network to automatically check news veracity in microblogs. They used tweets from Sina Weibo Chinese microblog and their proposed model called CPCV (Credibility Propagation with Conflicting Viewpoints) achieved an accuracy of 84% which was better than the accuracy by 4% compared with Jin et al. [61].

Janze et al. [62] built a model for investigational study based on blended methods of Elaboration Likelihood Model and User-Generated Content employing emotional, cognitive, behavioral, and visual indicators of the news shared on Facebook. The paper utilized fact-checked data of Facebook posts from the BuzzFeed News dataset and the predictive accuracy of the top-performing configurations was greater than 80 percent. The major limitation of the proposed work is that it only uses Facebook messages which are different from other social media platforms in the structure as well as at the function level. The paper pointed to getting their proposed detection system bypassed by any future advancement made in natural language generation using the findings suggested in the paper.

Wang et al. [63] introduced the LIAR dataset to support the building of computational and statistical approaches for spotting fake news. Further, the study used five baselines (SVM, CNNs, Bi-LSTMs, majority baseline, and regularised LR) to analyze patterns in surface-level linguistic data in the proposed dataset in order to spot fake news where the experiment results showed that CNNs (Convolutional Neural Networks) based model surpassed all models in performance, yielding an accuracy of 27 percent on the held-out test set.

Buntain et al. [64] built a classification model to determine whether a tweet in a thread on Twitter is real or false where Tweet Events from CREDBANK and PHEME datasets were provided as input data for training and testing the model. The proposed model performed better than Castillo et al. [65] with an accuracy of 61.81 percent as opposed to 66.93 percent accuracy using the PHEME dataset and 70.28 percent accuracy using the CREDBANK dataset.

Ruchansky et al. [40] created a hybrid deep learning-based model, named CSI (Capture, Score, and Integrate), so as to recognize false news with respect to the user and article by identifying three distinct characteristics (text, reaction, and source) of fake information. Despite the lacking of used related ground-truth labels, the CSI model achieved 89.2% accuracy on the Twitter dataset and 95.3% accuracy on the Weibo dataset.

Kotteti et al. [66] employed TF-IDF vectorization (in order to eliminate unnecessary features) and data imputation to fill in missing values for categorical and numerical features in a dataset. The paper used short statements from the LIAR dataset as input and 86 percent accuracy was obtained utilizing the supervised classifier.

Aldwairi et al. [9] offered a simple browser add-on that might be used to identify and filter out potential clickbait. The paper used news statements from social media sites as input during the experimental setup and used the logistic classifier to attain a 99.4 percent accuracy.

Pan et al. [67] combined methods of knowledge graphs and the TransE model to create the B-TransE model in order to effectively handle the issue of content-based spotting of fake news. They provided news articles as input to the proposed models. The (FML + TML-D4) version of the model achieved the following accuracies - (i). 0.89 using the poly-kernel function of the SVM (Support Vector Machine) classifier, (ii). 0.87 using the linear function of the SVM classifier, and (iii). 0.81 using the RBF function of the SVM classifier

Della et al. [68] developed a model, named HC-CB-3 (where HC stands for 'Harmonic boolean label Crowdsourcing', and CB stands for 'Content-Based'), to detect fake news automatically. The paper provided Facebook posts and their likes as input to the model and attained 81.7 percent accuracy.

Atodiressei et al. [69] created a service-based architecture for spotting fake users as well as fake news on Twitter. In order to build the system, the authors used Java as the programming language, Node.js for persistent storage, NoSql as the database, and MongoDB to store unstructured information, and they organized contents into two collections - users and tweets. The experiment results showed that the system discovered 88 similar tweets with a confidence score of 60.0. When the same tweet was retweeted by some other user, the system found 99 similar tweets with the same confidence score of 60.0.

Wang et al. [70] built an Adversarial Network-inspired event-invariant framework named EANN that contains (a). a feature extractor based on a multimodal technique, (b). a detector for locating false news, and (c). an event discriminator for spotting false news on an early basis. The authors provided Tweet events from real-time datasets, including Weibo and Twitter, as inputs to the system. It attained 71.5 percent accuracy on the Twitter dataset and 82.7 percent accuracy on the Weibo dataset.

Ajao et al. [71] created a hybrid model based on deep learning by employing LSTM (Long Short-Term Memory) and CNN (Convolutional Neural Network) models to automatically spot features inside posts and used Tweets as input for feeding the model in order to make fake news predictions. The model attained 82 percent accuracy without any prior understanding of the topics being covered. However, negative appreciation was observed against the LSTN model (plain-vanilla) due to a lack of necessary training examples.

Karimi et al. [72] designed a framework based on CNN and LSTM, called MMFD (Multi-source Multi-class Fake news Detection), by employing a combination of techniques including automated retrieval of features, fusion based on multiple sources, and spotting of fake news on an automated basis. The paper used statements from the LIAR dataset as input to the framework to predict fake news and it attained the maximum accuracy of 38.81 percent from several source combinations. The major limitation of the proposed framework is that it was found difficult to determine the fakeness of news from the profiles of writers and the short statements.

Wu et al. [73] created a TraceMiner model to detect false news using information from the diffusion network, even when the content details were absent. The authors used Tweets from Twitter as input to the model for predicting the fakeness of news. TraceMiner outperformed all baselines and its variations including TM(DeepWalk) and TM(LINE) with sample ratios ranging from 0.1 to 0.9.

Gupta et al. [74] treated a news article as a three-mode tensor structure called "News, User, Community" and latently embedded the news articles in order to simulate the "echo chambers" inside the social media network to construct a model, named CIMTDetect (Community Infused Matrix Tensor Detect), that can identify fake news. The proposed model attained maximum scores of precision, recall, and F1-score of 0.843 ± 0.270 , 1.000 ± 0.000 and 0.813 ± 0.078 on the BuzzFeed News dataset while it achieved high scores of 0.842 ± 0.191 precision, 0.933 ± 0.041 recall, and 0.818 ± 0.069 F1-score on the Politifact dataset.

Dong et al. [75] developed a hybrid attention model, called DUAL, for the prediction of fake news. The model was built using a bidirectional attention-based GRU (Gated Recurrent Units) for feature retrieval, which is then applied to produce an attention matrix for identifying false news. The authors used short statements from both LIAR and BuzzFeed News datasets as input to the proposed model and obtained 82.3 percent accuracy using the LIAR dataset as opposed to 83.8 percent accuracy using the BuzzFeed News dataset.

Guacho et al. [76] created a KNN graph of news items by utilizing embeddings of tensor-based news pieces to identify their similarities in embedding and latent space. Even though the proposed method only used 30 percent of the labels from a public dataset, it surpassed SVM classifier with 75.43 percent accuracy.

Helmstetter et al. [77] applied a weakly supervised approach for building a model that automatically gathered a large-scale noisy training dataset containing tweets. Further, the tweets were automatically labeled during data collection through their sources either as trustworthy or untrustworthy. The model, when evaluated, detected fake news with an F1 score up to 0.9 even though the dataset was highly noisy.

Liu et al. [78] constructed a classifier based on time series utilizing Convolution Neural Network and Recurrent Neural Network to detect false news by determining the changes of local as well as global kinds during the identification of users along the path of news propagation. The authors provided sharing cascades on Weibo, Twitter15, and Twitter16 datasets as input to the classifier and it surpassed all baselines with 85 percent accuracy using the Twitter datasets and 92 percent accuracy using the Sina Weibo datasets.

Shu et al. [79] measured users' sharing behaviors to identify group representatives' chance of sharing fake news, and performed a comparison among characteristics of their implicit and explicit profiles to identify fake news. The paper used statement claims from the Politifact and the Gossipcop dataset while testing the model for fake news prediction and achieved 96.6% accuracy using the Gossipcop dataset and 90.0% accuracy utilizing the PolitiFact dataset while comparing the effects of profile features as a whole versus only implicit or explicit features.

Reis et al. [80] applied supervised learning classifiers over a dataset of news articles with characteristics drawn from news sources, news sources' environments, and news content. Random forests and XGBoost classifiers outperformed all baselines in terms of AUC (Area Under the ROC Curve) and F1-score. The random forest classifier achieved an AUC of 0.85 ± 0.007 and an F1-score of 0.81 ± 0.008 . XGBoost classifier achieved an AUC of 0.86 ± 0.006 and an F1-score of 0.81 ± 0.011 .

Olivieri et al. [81] merged the textual characteristics with that of the task-generic characteristics (retrieved from Google's search engine results during crowdsourcing of missing values) for building a fake news detection model and used news articles from the PolitiFact website as input to the model. The experimental results demonstrated a significant improvement in F1-Score of 3 percent over the state-of-the-art techniques for a six-class task.

Bharadwaj et al. [82] employed semantic elements in conjunction with a number of machine-learning techniques to identify false news on the internet. The short statements

of real_or_fake dataset were used as input for the experimental setup. Bigrams outperformed a random forest classifier to achieve an accuracy of 95.66 percent.

Rasool et al. [83] utilized iterative learning and dataset relabeling to build a multilevel, multiclass technique for identifying fake news. The authors used short statements from the LIAR dataset as input to the model for fake news prediction. The proposed technique achieved an accuracy of 39.5%.

Traylor et al. [84] utilized Textblob, NLP, and SciPy Toolkits to build a system for the identification of fake news through influence mining (a cited attribution procedure) in a Bayesian classifier of machine learning. The precision of the proposed system in predicting if a heavily quoted article is likely to be fraudulent was 63.333 percent.

Shu et al. [85] built a FakeNewsTracker model using deep LSTMs (Long Short-Term Memories) where both the encoder and the decoder have two layers with 100 cells each. The authors provided short statements from the BuzzFeed News dataset as well as the Politifact dataset as input to the model for the fake news prediction. The proposed model achieved 67% accuracy on the PolitiFact dataset and 74.2% accuracy on the BuzzFeed News dataset.

Yang et al. [86] developed a probabilistic graphical model, called UFD (Unsupervised Fake News Detection), by employing the data of social engagement hierarchy to extract the opinions of social media users, along with the Gibbs method for sampling, for the purposes of identification of false news and determination of the credibility of users. The authors provided short statements from the LIAR dataset and the BuzzFeed News dataset as input to the model in order to detect fake news. The model achieved 67.9 percent accuracy using the BuzzFeed News dataset and 75.9 percent accuracy utilizing the LIAR dataset.

Kesarwani et al. [87] employed the K-Nearest Neighbor classifier and provided short statements from the BuzzFeed News dataset as input to the classifier in order to spot false news related to social media. The classifier achieved 79% classification accuracy.

Shu et al. [88] developed propagation networks, called STFNN and HPFN, and compared as well as contrasted the features of hierarchical propagation network and the structural/temporal/linguistic features of fake/real news in order to experimentally check their role in spotting fake news. The authors used short statements from the Politifact dataset and the Gossipcop dataset as input to models for fake news prediction. The combined STFNN-HPFN propagation network model attained 86.3 percent accuracy utilizing the Gossipcop dataset and 85.6 percent accuracy using the PolitiFact dataset.

Zhou et al. [89] developed a theory-driven model built using supervised learning classifiers to analyze news material at the lexical, syntactic, semantic, social, forensic psychology, and discourse levels to identify fake news. The paper used short statements of the two datasets. PolitiFact and BuzzFeed News, for experimental purposes. The proposed model achieved 50 to 60 percent accuracy using the BuzzFeed News dataset and 60 to 70 percent accuracy utilizing the PolitiFact dataset.

Lu et al. [90] created a model, called GCAN, based on the idea of an attention mechanism of neural network for the purpose of spotting fake tweets if the original tweet and the corresponding series of retweets without comments are provided as input. The authors used sharing cascades from Twitter15 and Twitter16 datasets as inputs to the model for fake news prediction. On the Twitter15 dataset, the GCAN model surpassed all the baselines with 87.67 percent accuracy utilizing the Twitter16 dataset.

Ni et al. [91] built an MVAN (Multi-View Attention Networks) model based on deep learning for the purpose of spotting fake news on an early basis. They merged the text semantic attention and the propagation structure attention in the model in order to simultaneously gather important hidden cues from the originating tweet's dissemination structure. The authors provided sharing cascades from Twitter15 and Twitter16 datasets as input to the model for

fake news prediction. MVAN model surpassed all baselines, obtaining 92.34 percent accuracy using the Twitter15 dataset and 93.65 percent accuracy utilizing the Twitter16 dataset.

Sahoo et al. [92] utilized LSTM (Long Short-Term Memory) analysis to create a system in the Chrome environment that can detect fake news on Facebook. The authors provided news articles from the FakeNewsNet dataset as input to the system for spotting fake news. The proposed system attained 99.4 percent accuracy.

Kaliyar et al. [93] developed FakeBERT, a BERT-based deep learning model, to identify fake news by combining numerous blocks of the one-layer Convolution Neural Network in parallel with myriad kernel sizes and filters. The proposed model had the benefit to deal with the ambiguity issue observed during natural language processing and it outperformed all baselines by 98.90% accuracy.

Dou et al. [94] developed the UPFD (User Preference-aware Fake news Detection) framework to spot fake news that used the combined approaches of graph modeling and news content to simultaneously collect myriad signals related to user preferences. News articles from the FakeNewsNet dataset were provided as input to the framework for fake news prediction. The UPFD achieved an accuracy of 84.62% and F1-score of 84.65% on the Politifact dataset. It achieved an accuracy of 97.23% and F1-score of 97.22% on the Gossipcop dataset.

Verma et al. [95] proposed a two-phase model, called WELFake (Word Embedding over Linguistic features for Fake news prediction), that provided machine learning-based classification of fake news. The dataset is preprocessed and the linguistic feature is used to validate news veracity during the first phase of the model, followed by the merging of Word Embedding with linguistic features followed by utilizing voting classification during the second phase. The news article from the well-curated WELFake dataset was used as input for the system during the prediction of fake news and it achieved 96.73% accuracy which is 1.31% better in comparison to the BERT-based model, and 4.25% better in comparison to the CNN-based model.

Ozbay and Alatas [96] developed a model using standard SSO (Salp Swarm Optimization) technique according to a nonlinear decrement coefficient and a varying inertia weight in order to obtain the highly optimum solution for spotting fake news. The short statements from real word datasets including LIAR, BuzzFeed News, ISOT, and Random political news were provided as input to the experimental setup. The system achieved 78.0 percent accuracy utilizing the LIAR dataset, 80.3 percent accuracy on the BuzzFeed News dataset, 85.9 percent accuracy using the ISOT fake news dataset, and 71.3 percent accuracy on the Random political news dataset.

Nasir et al. [97] suggested a hybrid model based on deep learning for detecting fake news that uses recurrent and convolutional neural networks. The real-time datasets including ISOT and FA-KES were used to validate the model and the hybrid model outperformed the non-hybrid model by 0.60 ± 0.007 accuracy on the FA-KES dataset, and by 0.99 ± 0.02 accuracy on the ISOT dataset.

Iwendi et al. [98] proposed thirty-nine text features using Deep Learning classifiers in order to identify fake news related to COVID-19. The authors built the dataset using news and other information from various social media sites. The proposed features detected fake news with 86.12% accuracy, a 20% improvement through the proposed features.

Seddari et al. [99] suggested a hybrid approach to integrate knowledge-based and linguistic elements to recognize false news. The short statements from the BuzzFeed News dataset were provided as input to the system for fake news prediction. The accuracy of the proposed system obtained was 94.4%.

Min et al. [100] proposed a model for spotting fake news, called PSIN (Post-User Interaction Network), that utilized the strategy of divide-and-conquer to effectively determine the

interactions of post-and-user, post-and-post, and user-and-user with respect to social context while preserving their implicit features. Further, an adversarial topic discriminator was employed for the topic-rationalist feature learning so as to enhance the capability of the system to generalize the newly introduced topics. The system is provided with short news statements as input from the existing multiple datasets like Politifact, Snopes, etc. The proposed PSIN model outperformed all the baselines by an average AUC (Area under the ROC curve) of 0.6367 and 0.3094 mean F1-score.

Galli et al. [101] developed a framework based on machine/deep learning techniques for spotting fake news while utilizing myriad feature arrangements discussed in the literature. The inputs from the real-world datasets including LIAR, FakeNewsNet, and PHEME were provided to the system for making predictions of fake news. The experiment was performed on various machine learning classifiers including SGD, Naive Bayes, Linear SVC, Random Forest, Logistic Regression, Nearest Neighbor, Decision Tree, Gradient Boost, Perceptron, and Passive Aggressive where the framework achieved the highest accuracy of 62.7% using Logistic Regression classifier on the LIAR dataset while it achieved the highest frequency of 59.6% using the Logistic Regression classifier using the Politifact dataset. The experiments were also performed using different deep learning classifiers including BERT (Bidirectional Encoder Representations from Transformers), C-HAN (Convolutional Hierarchical Attention Networks), BI-LSTM (Bidirectional Long-Short Term Memory), and CNN (Convolutional Neural Network). The framework achieved 61.9 percent accuracy using the BERT classifier on the LIAR dataset while it attained 58.8 percent accuracy using the BERT classifier on the Politifact dataset.

Sadeghi et al. [102] designed a method to spot false news using the NLI (Natural Language Inference) technique where the relevant and similar information that appeared in reliable news outlets are utilized as additional information in order to determine the truthfulness of a particular news article. The proposed method achieved 85.58% accuracy on the FNID-FakeNewsNet and 41.31% accuracy on the FNID-LIAR dataset.

Chi and Liao [103] proposed a model, named QA-AXDS, to automatically check the news veracity on social media while providing the necessary details about the results at the same time. The model contains key features including (a). it is data-driven; (b). it is supposed to be more scalable than the existing quantitative models; (c). it is supposed to be human-interactive and can automatically acquire knowledge at the human level; (d). it is supposed to provide greater interpretability and transparency than the existing machine learning-based models; and (e). its explanation method is supposed to increase the chances of improving the algorithms while working on the explanations of the raised issues. The authors used datasets including Twitter 2017, Twitter 2019, and Reddit for providing input to the system for fake news prediction. The experiments were performed using different classifiers including BranchLSTM, B.E.R.T., Xgboost, SVM (Support Vector Machine), LR (Logistic Regression), MLP (Multilayer Perceptron). The model achieved the highest accuracy of (a). 0.784 on the Twitter 2017 dataset using the BranchLSTM classifier; (b). 0.778 on the Twitter 2019 dataset using the BranchLSTM classifier; and (c). 0.929 on the Reddit dataset using the BranchLSTM and the LR classifiers.

Choudhury and Acharjee [104] developed a genetic algorithm-based model to spot fake news on social media. The experimental results showed that the model achieved 61 percent accuracy on the Liar dataset, 97 percent accuracy on the Fake Job Posting dataset, and 96 percent accuracy on the Fake News dataset using SVM (Support Vector Machine) classifier. All the papers reviewed so far in this section has been compared in Table 3.

Table 3 Comparison among the reviewed research articles

Algorithm	Authors	Strengths	Weaknesses	Performance
Decision-tree algorithm	Castillo et al. [10]	Utilizes a broad feature set	Limited to Twitter events; no comparison with other models	Accuracy: 86%
SVM classifier	Rubin et al. [58]	Incorporates Rhetorical Structure Theory and Vector Space Modeling; Outperforms humans in fake news detection	Accuracy (63%) lower than chance accuracy (56%)	63%
Multiclass logistic regression	Ferreira et al. [59]	Successfully classifies the stance of news	Limited to rumored claims and news articles from the emergent dataset	Accuracy: 73%
Credibility propagation network	Jin et al. [41]	Utilizes conflicting social opinions; Outperforms a previous model by 4%	Specific to analyzing credibility in microblogs from Sina Weibo	Accuracy: 84%
Investigational study	Janze et al. [62]	Considers emotional, cognitive, behavioral, and visual indicators of news shared on Facebook	Relies solely on Facebook messages; may not generalize to other social media platforms	Accuracy: >80%
Baseline models on the LIAR dataset	Wang et al. [63]	Introduces the LIAR dataset; CNN-based model outperforms other baselines	CNN-based model achieves relatively low accuracy of 27% on the held-out test set	Accuracy: 27%
Classification model	Buntain et al. [64]	Outperformed a previous model by Castillo et al. [65]	Limited to Twitter threads; limited comparison with other models	Accuracy: 61.81%
Hybrid deep learning model	Ruchansky et al. [40]	Recognizes false news based on text, reaction, and source; Achieved high accuracy on Twitter and Weibo datasets	Lacking related ground-truth labels; performance on datasets other than Twitter and Weibo not mentioned	Accuracy: 89.2%, Weibo Accuracy: 95.3%

Table 3 continued

Algorithm	Authors	Strengths	Weaknesses	Performance
TF-IDF vectorization and data imputation	Kotteti et al. [66]	Uses TF-IDF vectorization and data imputation; Achieved relatively high accuracy	Limited to short statements from the LIAR dataset; no comparison with other models	Accuracy: 86%
Browser add-on	Aldwairi et al. [9]	Simple browser add-on with high accuracy in identifying and filtering potential clickbait	Relies on news statements from social media sites; limited to identifying clickbait; no comparison with other models	Accuracy: 99.4%
Knowledge graph and transE model	Pan et al. [67]	Uses knowledge graphs and the B-TransE model for content-based spotting of fake news	Performance measured by accuracy with different SVM classifiers using various functions; no direct accuracy comparison with other models	SVM Accuracy: 0.81–0.89
HC-CB-3 Model	Della et al. [68]	HC-CB-3 model detects fake news automatically based on Facebook posts and their likes	Limited to Facebook posts and likes; no comparison with other models	Accuracy: 81.7%
Service-Based Architecture	Atodireset et al. [69]	Architecture for spotting fake users and fake news on Twitter; Achieved successful discovery of similar tweets with high confidence scores	Limited to Twitter; performance measured in terms of finding similar tweets; no direct accuracy comparison with other models	Confidence Score: 60.0
Adversarial network-inspired framework	Wang et al. [70]	EANN framework for event-invariant false news detection; Achieved high accuracy on Twitter and Weibo datasets	Performance measured on Twitter and Weibo datasets; no direct accuracy comparison with other models	Twitter Accuracy: 71.5% Weibo Accuracy: 82.7%
Hybrid model based on LSTM and CNN	Ajao et al. [71]	Hybrid model employing LSTM and CNN for fake news detection in tweets; Achieved high accuracy without prior topic understanding	Negative performance observed for the LSTN model; no comparison with other models	Accuracy: 8.2%

Table 3 continued

Algorithm	Authors	Strengths	Weaknesses	Performance
MMFD (Multi-source multi-class fake news detection)	Karimi et al. [72]	Employed CNN and LSTM Fusion based on multiple sources; Automated feature retrieval	Difficulty determining fakeness from writer profiles and short statements	Maximum accuracy of 38.81%
TraceMiner	Wu et al. [73]	Detection using diffusion network information; Effective when content details are absent	NA	Outperformed baselines, sample ratios ranging from 0.1 to 0.9
CIMTDetect	Gupta et al. [74]	Modeled news articles as a tensor structure; Considered user and community influence	NA	F1 score of 0.813, 0.818 for BuzzFeed News, Politifact respectively
DUAL	Dong et al. [75]	Hybrid attention model using bidirectional attention-based GRU; Attention matrix for identifying false news	NA	Accuracy: LIAR - 82.3%, BuzzFeed News - 83.8%
KNN graph	Guacho et al. [76]	Utilized tensor-based news embeddings; Identified similarities in embedding and latent space	Relied on limited labels from a public dataset	Accuracy: 75.43%
Weakly supervised model	Helmstetter et al. [77]	Gathered a large-scale noisy training dataset containing tweets	Highly noisy dataset	Detected fake news with F1 score up to 0.9
Time series classifier	Liu et al. [78]	Used time series, CNN, and RNN; Identified changes in user characteristics during news propagation	NA	Accuracy: Twitter datasets - 85%, Sina Weibo datasets - 92%
Profile-based detection	Shu et al. [79]	Measured users' sharing behaviors; Compared implicit and explicit profile characteristics	NA	Accuracy: Gossipcop dataset - 96.6%, Politifact dataset - 90.0%

Table 3 continued

Algorithm	Authors	Strengths	Weaknesses	Performance
Random forest and XGBoost classifiers	Reis et al. [80]	Applied supervised learning classifiers; Utilized characteristics from news sources, environments, and content	NA	Random Forest: AUC 0.85 ± 0.007, F1-score 0.81 ± 0.008, XGBoost: AUC 0.86 ± 0.006, F1-score 0.81 ± 0.011
Task-generic characteristics	Olivieri et al. [81]	Merged textual and task-generic characteristics; Improved F1-Score for a six-class task	NA	Improved F1-Score by 3% compared to state-of-the-art techniques
Semantic elements and machine learning	Bharadwaj et al. [82]	Utilized semantic elements and machine learning techniques; Bigrams outperformed random forest classifier	NA	Accuracy: 95.66%
Iterative learning and dataset relabeling	Rasool et al. [83]	Multilevel, multiclass technique; Used iterative learning and dataset relabeling	NA	Accuracy: 39.5%
FakeNewsTracker using deep LSTMs	Shu et al. [85]	Deep LSTMs for fake news prediction; Achieved high accuracy on BuzzFeed News and PolitiFact datasets	NA	PolitiFact dataset: 67% accuracy; BuzzFeed News dataset: 74.2% accuracy
Unsupervised Fake News Detection (UFD)	Yang et al. [86]	Probabilistic graphical model using social engagement hierarchy; Determination of credibility of users	NA	BuzzFeed news dataset: 67.9% accuracy; LIAR dataset: 75.9% accuracy
K-nearest neighbor classifier	Kesarwani et al. [87]	Employed K-Nearest Neighbor classifier; Spotting false news related to social media	NA	Classification accuracy: 79%
Propagation networks (STFN, HFFN)	Shu et al. [88]	Experimental comparison of hierarchical propagation network features; Considered structural, temporal, and linguistic features	NA	Gossipcop dataset: 86.3% accuracy; PolitiFact dataset: 85.6% accuracy

Table 3 continued

Algorithm	Authors	Strengths	Weaknesses	Performance
Theory-driven model with supervised learning classifiers	Zhou et al. [89]	Analyzed news material at multiple levels; Considered lexical, syntactic, semantic, social, forensic psychology, and discourse features	NA	BuzzFeed News dataset: 50–60% accuracy; PolitiFact dataset: 60–70% accuracy
GCAN	Lu et al. [90]	Utilized attention mechanism in a neural network; Achieved high accuracy on Twitter datasets	Accuracy on Twitter16: NA	Accuracy on Twitter15: 87.67%
MVAN (Multi-View Attention Networks)	Ni et al. [91]	Merged text semantic attention and propagation structure attention; Outperformed all baselines on Twitter datasets	NA	Accuracy on Twitter15: 92.34%; Accuracy on Twitter16: 93.65%
LSTM (Long Short-Term Memory) analysis	Sahoo et al. [92]	Utilized LSTM analysis to detect fake news on Facebook. Achieved high accuracy on the FakeNewsNet dataset	NA	Accuracy: 99.4%
FakeBERT (BERT-based deep learning model)	Kaliyar et al. [93]	Combined BERT-based model with Convolutional Neural Network; Addressed ambiguity issues in natural language processing; Outperformed all baselines by 98.90% accuracy	NA	Accuracy: 98.90%
UPFD (User Preference-aware Fake news Detection)	Dou et al. [94]	Utilized graph modeling and news content approaches; Simultaneously collected signals related to user preferences; Achieved high accuracy on PolitiFact and Gossipcop datasets	NA	PolitiFact Accuracy: 84.62%, F1-score: 84.65%; Gossipcop Accuracy: 97.23%, F1-score: 97.22%
WELFake (Word Embedding over Linguistic features)	Verma et al. [95]	Two-phase model combining linguistic features and word embeddings; Outperformed BERT-based and CNN-based models	NA	Accuracy: 96.73%
SSO (Salp Swarm Optimization)	Ozbay and Alatas [96]	Utilized a nonlinear decrement coefficient and varying inertia weight; Achieved relatively high accuracy on LIAR, Buzzfeed News, ISOT, and Random political news datasets	NA	LIAR Accuracy: 78.0%; Buzzfeed News Accuracy: 80.3%; ISOT Accuracy: 85.9%; Random political news Accuracy: 71.3%

Table 3 continued

Algorithm	Authors	Strengths	Weaknesses	Performance
Hybrid model (Recurrent and Convolutional Networks)	Nasir et al. [97]	Combined recurrent and convolutional neural networks in a hybrid model; Outperformed non-hybrid model on FA-KES and ISOT datasets	NA	FA-KES Accuracy improvement: 0.60 ± 0.007 ; ISOT Accuracy improvement: 0.99 ± 0.02
Deep Learning classifiers	Iwendi et al. [98]	Proposed thirty-nine text features for identifying COVID-19 related fake news; Achieved improved accuracy compared to previous models	NA	Accuracy: 86.12% (20% improvement compared to previous models)
Hybrid approach (Knowledge-based and Linguistic)	Seddari et al. [99]	Integrated knowledge-based and linguistic elements to recognize false news; Achieved high accuracy on the BuzzFeed News dataset	NA	Accuracy: 94.4%
PSIN (Post-User Interaction Network)	Min et al. [100]	Utilized divide-and-conquer strategy and adversarial topic discriminator. Outperformed baselines in terms of AUC and F1-score	NA	Average AUC: 0.6367; Mean F1-score: 0.3094
Framework based on machine/deep learning techniques	Galli et al. [101]	Developed a framework using various feature arrangements and classifiers; Achieved highest accuracy with Logistic Regression classifier on the LIAR dataset and the Politifact dataset	NA	Accuracy of Logistic Regression on LIAR, Politifact are 62.7%, 59.6% respectively; Accuracy of BERT on LIAR, Politifact are 61.9%, 58.8% respectively
NLI (Natural Language Inference) technique	Sadeghi et al. [102]	Used NLI technique with reliable news outlets as additional information; Achieved high accuracy on the FNID-FakeNewsNet and FNID-LIAR datasets	NA	FNID-FakeNewsNet Accuracy: 85.58%; FNID-LIAR Accuracy: 41.31%
QA-AXDS model	Chi and Liao [103]	Data-driven, scalable, human-interactive model; Provides interpretability and transparency	NA	Twitter 2017 Accuracy: 0.784; Twitter 2019 Accuracy: 0.778; Reddit Accuracy: 0.929
Genetic algorithm-based model	Choudhury and Acharjee [104]	Utilizes genetic algorithms; Achieved high accuracy on Liar, Fake Job Posting, and Fake News datasets	NA	Liar Accuracy: 61% <; Fake Job Posting Accuracy: 97%; Fake News Accuracy: 96%

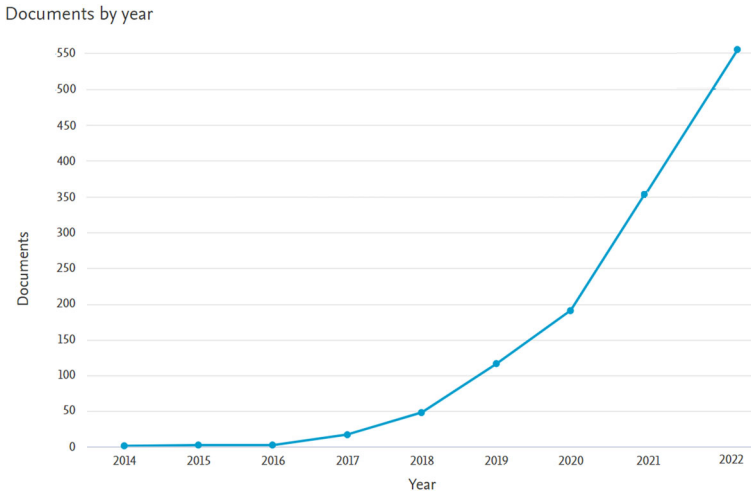


Fig. 5 Yearly publication of documents on the detection of social media related fake news

5 Results and discussion

The detection of social media related fake news has been one of the most anticipated fields of research in recent years. Following the 2016 US Presidential election, research related to the identification of fake news on social media has dramatically increased. This is demonstrated by Figs. 5, 6, and 7, which were collected from Scopus eSS¹ when searching for the phrase "Detection of Social Media Related Fake News". The amount of documents released annually from 2014 to 2022 is shown in Fig. 5. Additionally, Fig. 6 outlines the various document kinds that were released till December 2022. Surprisingly, conference papers have produced the most empirical/experimental research in the field of false news detection with 64.2%, while editorial and retracted have contributed the least with 0.1%.

Figure 7 provides a pie chart explaining the key domains for paper publications based on the spotting of social media related fake news. The pie chart simply indicates that the Computer Science field has the maximum publications in the area of spotting fake news at 42.8%, while the Energy field has contributed the least with 1.4%. The presented analysis further emphasizes that social media fake news detection research is a diverse area of study rather than being solely confined to the computer science and engineering field.

5.1 Fake news detection challenges

According to Zhang et al. [55], fake news has three characteristics represented by 3 Vs: Volume, Veracity, and Velocity. Following a thorough analysis of papers on the detection of social media-related fake news, we came to the conclusion that the two additional characteristics denoted by 2 Vs (Variety and Valid Dataset) besides the 3 Vs as described by Zhang et al. together constitute five Vs as significant challenges to developing a reliable and effective fake news detection system as shown in Fig. 8.

¹ <https://www.scopus.com>

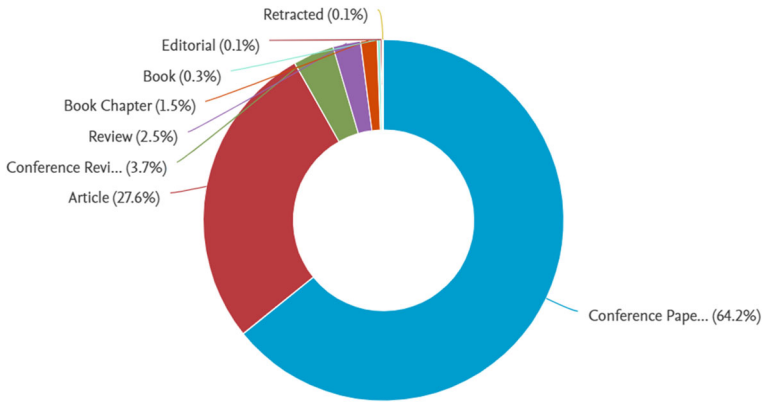


Fig. 6 Published document types in relation to the identification of social media related fake news

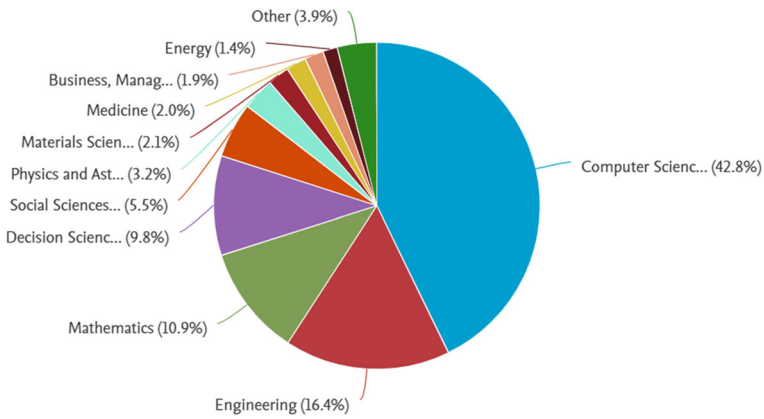


Fig. 7 Key domains for paper publication related to the spotting of social media based fake news

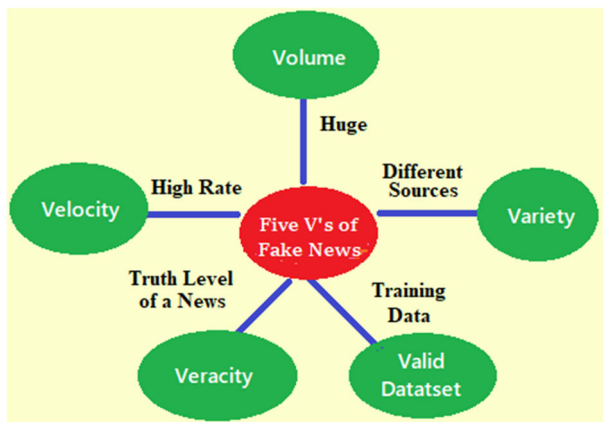


Fig. 8 Five V's related to fake news as key challenges for its detection

- **Volume:** It is not possible to properly fact-check everything using human experts since fake news is created quickly and widely [6]. This also necessitates developing early-detection technologies for fake news so as to restrict its propagation online [78].
- **Veracity:** It can be challenging to distinguish between real and false news because fake news is intentionally produced to mislead readers and mimic legitimate news sources. [6, 7, 55].
- **Velocity:** The lifespan of fake news publishers is short [2]. It is considerably more challenging to identify fake news because it is disseminated on social media in real-time. It's challenging to determine the number of internet users who are actively interacting with a specific piece of viral news [55].
- **Variety:** Fake news affects many different elements of people's lives and can be classified in a variety of interconnected ways, including hoax, disinformation, fake reviews, satire news, rumors, propaganda, click-bait, and so forth [55].
- **Valid dataset:** Very few training datasets for the purpose of spotting fake news have been provided by the research community, which generally do not have complete information with respect to misleading news, just because of restrictions made by social media sites on the collection of data related to public domain [105]. A valid dataset is crucial for checking news veracity. Some of the widely used datasets for carrying out the automatic spotting of social media related fake news are described in the subsection–5.2

5.2 Dataset

A benchmark dataset is crucial for the computational work being done on fake news identification on social media. The commonly utilized datasets in the publications analyzed during the current review study include Politifact,² LIAR,³ FakeNewsNet,⁴ BS Detector,⁵ Twitter,⁶ BuzzFeed News,⁷ Weibo,⁸ PHEME,⁹ and CREDBANK.¹⁰ Table 4 summarizes the quality, advantages, and disadvantages of these datasets used in the field of spotting fake news and how well-suited they are for various fake news detection tasks. Public fake-news databases come in three different varieties: *claims* (they can consist of one or more sentences to provide information that needs to be validated, e.g. POLITIFACT), *complete articles* (single piece of information consisting of interconnected claims of several kinds, e.g. BS DETECTOR and FAKENEWSNET), and *SNS (Social Networking Services) data* (despite being same in size as that of claims, it also contains a lot of non-text data, such as information from BuzzFeed News, PHEME, and Credbank, as well as structured data from accounts and posts).

There are certain datasets that have been developed in recent years covering the fake news events of either multiple regions or multiple languages. For instance, MultiFC Dataset [106] (it covers online fake news events in English, Spanish, Arabic, and Russian languages), BharatFakeNewsKosh dataset [107] (it is a dataset specifically focused on fake news in India

² <https://www.kaggle.com/general/232077>

³ <https://paperswithcode.com/dataset/liar>

⁴ <https://www.kaggle.com/datasets/mdepak/fakenewsnet>

⁵ <https://github.com/thiagovas/bs-detector-dataset/>

⁶ <https://www.kaggle.com/general/35739>

⁷ <https://github.com/BuzzFeedNews/>

⁸ <https://github.com/ww-rm/weibo-rmdt>

⁹ <https://www.kaggle.com/datasets/usharengaraju/pheme-dataset>

¹⁰ <https://github.com/compsocial/CREDBANK-data>

Table 4 Some popular fake news related datasets

Dataset	Size	Composition	Platform	Purpose	Quality	Advantages	Disadvantages
PolitiFact	488 Claims	Multiple	Unknown	Spotting False News	Moderate	Contains political claims for evaluating political news	Limited information about the platform and collection
LIAR	10,000 (approx.)	Claims	Multiple	Spotting False News	Moderate	Large dataset with diverse claims	Limited information about the platform and collection
FakeNewsNet	1,000 (approx.)	Complete articles	Twitter	Spotting False News	Moderate	Includes complete articles and metadata from Twitter	Relatively small dataset
BS Detector	10,000 (approx.)	Complete articles	Multiple	Spotting False News	Moderate	Large dataset with complete articles from multiple sources	Limited information about the platforms and collection
Twitter	1,000 (approx.)	SNS data	Twitter	Spot False News, Rumor	Moderate	Real-time data capturing and social network interactions	Limited dataset size and platform-specific biases
BuzzFeed News	2,282	SNS data	Facebook	Spotting False News	Moderate	Dataset from a popular social media platform	Relatively small dataset
Weibo	4,700	SNS data	Sina Weibo	Spot False News, Rumor	Moderate	Dataset from a popular Chinese social media platform	Relatively small dataset
PHEME	6,500	SNS data	Multiple	Spot False News, Rumor	Moderate	Covers rumor propagation on Twitter during real events	Limited to specific events and time period
CREDBANK	60 million	SNS data	Twitter	Credibility Assessment	High	Large-scale dataset for assessing credibility on Twitter	Dataset primarily focused on credibility, not fake news

and covers nine Indian languages including Assamese, Bangla, English, Gujarati, Hindi, Malayalam, Odia, Tamil, and Telugu), and Arabic Fake News Dataset (AFND) [108] which comprises of Arabic news articles collected from various Arabic news websites.

5.3 Research gaps and future directions

The following research gaps have been identified as a result of the comprehensive review conducted in this paper. These gaps are important areas that need to be addressed in future research on the automatic detection of social media-related fake news:

1. **Real-time detection:** One major research gap is the need for methods that can reliably spot false news at an early stage in real-time. Developing a real-time detecting system for fake news is crucial to prevent the propagation of misinformation and enable timely intervention.
2. **Global benchmark dataset:** Fake news is a global issue, and each country faces unique challenges in addressing it. To tackle fake news on a global level, it is necessary to construct benchmark datasets that cover local and national instances of fake news. Combining these datasets from different nations can contribute to the development of a comprehensive global benchmark dataset for detecting fake news.
3. **Fine-grained classification:** While many current datasets use binary labels to classify news as true or false, there is a need for more precise classification methods that consider the intent behind the news. Differentiating between true fake news and related content such as opinion pieces and satirical news can provide more nuanced insights and improve the effectiveness of detection models.
4. **Alternative approaches:** In addition to machine/deep learning methods, there is a need to explore other reliable and efficient ways to identify false news. Research should focus on investigating alternative approaches and techniques that can complement existing methods and enhance the accuracy of fake news detection.
5. **Application in related areas:** Fake news identification has implications beyond its direct impact on spotting false news. Research in areas such as locating bots/spammers, detecting clickbait, identifying stance, and addressing rumors is vital to address the broader challenges associated with misinformation. Future work in these areas will contribute to a comprehensive understanding and mitigation of the difficulties related to fake news.
6. **Feature reduction:** Exploring techniques to reduce the size of feature vectors and effectively handle large amounts of fake news data is an important research direction. Developing efficient feature selection or extraction methods can improve the scalability and performance of fake news detection models.
7. **Visual content analysis:** There is a lack of trustworthy research on extracting features from visual entities like photographs and videos to develop systems for spotting fake news. Investigating the use of visual content analysis and integrating it into detection models can provide valuable cues and enhance the overall accuracy and effectiveness of fake news detection.
8. **Attention mechanism:** The attention mechanism, a technique that focuses on relevant parts of the input, shows promise in improving the accuracy of models for spotting false news. Further research in this area can explore the application of attention mechanisms to fake news detection models and investigate its impact on performance.
9. **Unraveling belief in fake news:** One of the potential research gap could be the need for a deeper understanding of the complex dynamics influencing belief in fake news.

This includes the interplay between media ecosystems, cognitive factors, and sociological influences. [109] highlights the importance of investigating the impact of information agendas, cognitive biases, emotional factors, and sociological factors such as echo chambers. By unraveling these complex dynamics, the research aims to address the challenges posed by the belief in fake news in contemporary society.

10. **Enhanced fake news analysis through advanced pattern-mining system:** There is a need for the development of advanced pattern-mining frameworks that can effectively analyze fake news data. Current techniques often struggle to provide actionable insights into the structure of fake news. The proposed DMRM-FNA (DMRM-Fake News Analysis) framework addresses this challenge by extracting patterns from fake news and comparing them with real news data [110]. Future research can focus on further improving the efficiency and accuracy of pattern mining techniques for analyzing large-scale social network data.
11. **Fake news propagation (epidemiological and non-epidemiological perspectives):** Understanding the dynamics of fake news propagation is crucial in combating misinformation. This research highlights the exploration of both epidemiological and non-epidemiological models to capture the spread of fake news [111]. Future studies can delve deeper into these models, including their extensions and modifications, to gain insights into real-world scenarios. This line of research enables a better understanding of the complex mechanisms behind fake news propagation.
12. **Temporal focus estimation in news articles:** Estimating the temporal focus of news articles plays a significant role in combating fake news and providing relevant contextual information. Ahmed et al. [112] combines co-training, lexicon expansion, and semi-supervised learning techniques to identify temporal expressions and classify articles based on criteria such as "what," "when," "where," and "who". Further research can focus on improving the accuracy and robustness of temporal focus estimation, exploring semantic sentence segmentation methods, and addressing challenges related to network instability.
13. **Leveraging lexicon expansion for event detection:** Incorporating lexicon expansion techniques can enhance the accuracy of event detection algorithms in news retrieval. This research demonstrates the benefits of WordNet-based lexicon expansion, contextual representations, and semi-supervised learning for training classifiers [113]. Future studies can explore further enhancements to the approach, such as incorporating additional contextual features, optimizing the matching classifiers, and evaluating the performance on larger datasets. The research contributes to improving temporal analysis in news article retrieval and provides a comprehensive framework for accurate time estimation.
14. **Transformer-based approach for fake news detection:** The utilization of Transformer-based models for fake news detection shows promise in improving classification accuracy. Raza et al. [114] proposes a framework that incorporates a Transformer-based encoder-decoder model and incorporates various features from news content and social contexts. Future research can explore the application of Transformer-based approaches in fake news detection, further refine the model architecture, and investigate techniques to address the challenge of limited labeled data.

By addressing these research gaps and exploring the future directions mentioned, the field of automatic spotting of social media-related fake news can make significant advancements. These efforts will contribute to the development of more robust and effective methods for identifying and combating fake news, enabling a more reliable and trustworthy information environment. The potential benefits include improved early detection and prevention of fake

news propagation, enhanced understanding of fake news dynamics, accurate temporal focus estimation, and more accurate and efficient fake news detection models.

6 Conclusion

This paper provides a comprehensive review of 60 published papers based on the automatic detection of social media related fake news. All the papers were selected between the time frame of the year 2011 to 2022 according to the PRISM research methodology. The summary of each reviewed article was presented while highlighting (a). key techniques adopted based on machine/deep learning, (b). datasets used for experimental setups, and (c). performance metrics of each model or technique. The major research findings of the review were discussed along with the key challenges faced during the building of an automatic detection system for identifying fake news. Finally, the major research gaps along with the future scope of automatic detection of social media related fake news were discussed.

Funding The authors received no specific funding for this work.

Data Availability Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

References

1. Kietzmann JH, Hermkens K, McCarthy IP, Silvestre BS (2011) Social media get serious! understanding the functional building blocks of social media. *Bus Horiz* 54(3):241–251
2. Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. *J Econ Perspect* 31(2):211–36
3. Perrin A (2015) Social media usage. *Pew Research Center* 125:52–68
4. Howell L et al (2013) Digital wildfires in a hyperconnected world. *WEF Report* 3(2013):15–94
5. Sailunaz K, Kawash J, Alhaji R (2022) Tweet and user validation with supervised feature ranking and rumor classification. *Multimed Tools Appl* 81(22):31907–31927
6. Shu K, Sliva A, Wang S, Tang J, Liu H (2017) Fake news detection on social media: a data mining perspective. *ACM SIGKDD Explorations Newsletter* 19(1):22–36
7. Shao C, Ciampaglia GL, Varol O, Yang K-C, Flammini A, Menczer F (2018) The spread of low-credibility content by social bots. *Nat Commun* 9(1):1–9
8. Tandoc EC Jr, Lim ZW, Ling R (2018) Defining fake news a typology of scholarly definitions. *Digit Journal* 6(2):137–153
9. Aldwairi M, Alwahedi A (2018) Detecting fake news in social media networks. *Procedia Comput Sci* 141:215–222
10. Castillo C, Mendoza M, Poblete B (2011) Information credibility on twitter. In: *Proceedings of the 20th international conference on world wide web* pp 675–684
11. Niklewicz K (2017) Weeding out fake news: an approach to social media regulation. *European View* 16(2):335–335
12. Pierri F, Ceri S (2019) False news on social media: a data-driven survey. *ACM Sigmod Record* 48(2):18–27
13. Fleming C, O'Carroll J (2010) The art of the hoax. *Parallax* 16(4):45–59
14. Wang C-C (2020) Fake news and related concepts: definitions and recent research development. *Contemp Manag Res* 16(3):145–174
15. Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. *Sci* 359(6380):1146–1151

16. Volkova S, Shaffer K, Jang JY, Hodas N (2017) Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter. In: Proceedings of the 55th annual meeting of the association for computational linguistics, (vol 2: Short Papers), pp 647–653
17. Chen Y, Conroy N.J, Rubin VL (2015) Misleading online content: recognizing clickbait as false news. In: Proceedings of the 2015 ACM on workshop on multimodal deception detection, pp 15–19
18. Alkhodair SA, Ding SH, Fung BC, Liu J (2020) Detecting breaking news rumors of emerging topics in social media. *Inf Process Manag* 57(2):102018
19. Shelke S, Attar V (2022) Rumor detection in social network based on user content and lexical features. *Multimed Tools Appl* 81(12):17347–17368
20. Rubin VL, Chen Y, Conroy NK (2015) Deception detection for news: three types of fakes. *Proc Assoc Inf Sci Technol* 52(1):1–4
21. Elhadad MK, Li KF, Gebali F (2019) Fake news detection on social media: a systematic survey. In: 2019 IEEE Pacific rim conference on communications computers and signal processing (PACRIM). IEEE, pp 1–8
22. Wardle C, Derakhshan H et al (2018) Thinking about information disorder: formats of misinformation disinformation and mal-information. Ireton Cherylyn; Posetti Julie. *Journalism fake news & disinformation*. Paris: Unesco, pp 43–54
23. Kshetri N, Voas J (2017) The economics of fake news. *IT Professional* 19(6):8–12
24. Rastogi S, Bansal D (2022) Disinformation detection on social media: an integrated approach. *Multimed Tools Appl* 81(28):40675–40707
25. Kucharski A (2016) Study epidemiology of fake news. *Nature* 540(7634):525–525
26. Upadhyay R, Pasi G, Viviani M (2022) Vec4cred: a model for health misinformation detection in web pages. *Multimed Tools Appl* pp 1–20
27. Zhou X, Zafarani R (2020) A survey of fake news: fundamental theories detection methods and opportunities. *ACM Comput Surv* 53(5):1–40
28. Rousidis D, Koukaras P, Tjortjis C (2020) Social media prediction: a literature review. *Multimed Tools Appl* 79(9–10):6279–6311
29. Howard PN, Kollanyi B (2016) Bots# strongerin and# brexit: computational propaganda during the uk-eu referendum. Available at SSRN 2798311
30. Rapoza K (2017) Can fake news impact the stock market. *Forbes News*
31. Ferrara E, Varol O, Davis C, Menczer F, Flammini A (2016) The rise of social bots. *Commun ACM* 59(7):96–104
32. How fake news in the middle east is a powder keg waiting to blow. *Independent*
33. Spohr D (2017) Fake news and ideological polarization: filter bubbles and selective exposure on social media. *Bus Inf Rev* 34(3):150–160
34. Shayan Sardarizadeh OR (2022) Ukraine invasion: false claims the war is a hoax go viral. *BBC Monitoring*
35. Salve P (2020) Manipulative fake news on the rise in india under lockdown: study. *indiaspend.com*
36. PTI (2022) Govt orders blocking of 18 indian four pakistani youtube-based news channels. *time-sofindia.indiatimes.com*
37. Jin Z, Cao J, Zhang Y, Zhou J, Tian Q (2016) Novel visual and statistical image features for microblogs news verification. *IEEE Trans Multimedia* 19(3):598–608
38. Kwon S, Cha M, Jung K, Chen W, Wang Y (2013) Prominent features of rumor propagation in online social media. In: 2013 IEEE 13th international conference on data mining. IEEE, pp 1103–1108
39. Ma J, Gao W, Wei Z, Lu Y, Wong K-F (2015) Detect rumors using time series of social context information on microblogging websites. In: Proceedings of the 24th ACM international on conference on information and knowledge management, pp 1751–1754
40. Ruchansky N, Seo S, Liu Y (2017) 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
41. Jin Z, Cao J, Zhang Y, Luo J (2016) News verification by exploiting conflicting social viewpoints in microblogs. In: Proceedings of the AAAI conference on artificial intelligence, vol 30
42. Assiroj P, Hidayanto AN, Prabowo H, Warnars HLHS et al (2018) Hoax news detection on social media: a survey. In: 2018 Indonesian association for pattern recognition international conference (INAPR). IEEE, pp 186–191
43. Vlachos A, Riedel S (2014) Fact checking: task definition and dataset construction. In: Proceedings of the ACL 2014 workshop on language technologies and computational social science, pp 18–22
44. Cohen S, Hamilton JT, Turner F (2011) Computational journalism. *Commun ACM* 54(10):66–71
45. Rubin VL, Lukoianova T (2015) Truth and deception at the rhetorical structure level. *J Assoc Inf Sci Technol* 66(5):905–917

46. Atanasova P, Nakov P, Márquez L, Barrón-Cedeño A, Karadzhov G, Mihaylova T, Mohtarami M, Glass J (2019) Automatic fact-checking using context and discourse information. *J Data Inf Qual* 11(3):1–27
47. Mohammad SM, Sobhani P, Kiritchenko S (2017) Stance and sentiment in tweets. *ACM Trans Internet Technol* 17(3):1–23
48. Moher D, Liberati A, Tetzlaff J, Altman DG, Altman D, Antes G, Atkins D, Barbour V, Barrowman N, Berlin JA et al (2009) Preferred reporting items for systematic reviews and meta-analyses: the prisma statement (chinese edition). *J Chin Integr Med* 7(9):889–896
49. Özgöbek Ö, Gulla JA (2017) Towards an understanding of fake news. In: CEUR workshop proceedings, vol 354
50. Meinert J, Mirbabaie M, Dungs S, Aker A (2018) Is it really fake—towards an understanding of fake news in social media communication. In: International conference on social computing and social media. Springer, pp 484–497
51. Kumar S, Shah N (2018) False information on web and social media: a survey. [arXiv:1804.08559](https://arxiv.org/abs/1804.08559)
52. Parikh S.B, Atrey PK (2018) Media-rich fake news detection: a survey. In: 2018 IEEE conference on multimedia information processing and retrieval (MIPR). IEEE, pp 436–441
53. Sharma K, Qian F, Jiang H, Ruchansky N, Zhang M, Liu Y (2019) Combating fake news: a survey on identification and mitigation techniques. *ACM Trans Intell Syst Technol* 10(3):1–42
54. Cardoso Durier da Silva F, Vieira R, Garcia AC (2019) Can machines learn to detect fake news a survey focused on social media. In: Proceedings of the 52nd Hawaii international conference on system sciences
55. Zhang X, Ghorbani AA (2020) An overview of online fake news: characterization detection and discussion. *Inf Process Manage* 57(2):102025
56. Mridha MF, Keya AJ, Hamid MA, Monowar MM, Rahman MS (2021) A comprehensive review on fake news detection with deep learning. *IEEE Access* 9:156151–156170
57. Bani-Hani A, Adedugbe O, Benkhelifa E, Majdalawieh M (2022) Fandet semantic model: an owl ontology for context-based fake news detection on social media. *Combating fake news with computational intelligence techniques*, pp 91–125
58. Rubin VL, Conroy NJ, Chen Y (2015) Towards news verification: deception detection methods for news discourse. In: Hawaii international conference on system sciences, pp 5–8
59. Ferreira W, Vlachos A (2016) Emergent: a novel data-set for stance classification. In: Proceedings of the 2016 conference of the north American chapter of the association for computational linguistics: human language technologies. ACL
60. Magnini B, Zanolini R, Dagan I, Eichler K, Neumann G, Noh T-G, Pado S, Stern A, Levy O (2014) The excitement open platform for textual inferences. In: Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations, pp 43–48
61. Jin Z, Cao J, Jiang Y-G, Zhang Y (2014) News credibility evaluation on microblog with a hierarchical propagation model. In: 2014 IEEE international conference on data mining. IEEE, pp 230–239
62. Janze C, Risius M (2017) Automatic detection of fake news on social media platforms
63. Wang WY (2017) liar liar pants on fire: a new benchmark dataset for fake news detection. In: Proceedings of the 55th annual meeting of the association for computational linguistics. Association for Computational Linguistics, (vol 2: Short Papers), pp 422–426. <https://doi.org/10.18653/v1/P17-2067>. <https://aclanthology.org/P17-2067>
64. Buntain C, Golbeck J (2017) Automatically identifying fake news in popular twitter threads. In: 2017 IEEE international conference on smart cloud (SmartCloud). IEEE, pp 208–215
65. Castillo C, Mendoza M, Poblete B (2013) Predicting information credibility in time-sensitive social media. *Internet Res*
66. Kotteti CMM, Dong X, Li N, Qian L (2018) Fake news detection enhancement with data imputation. In: 2018 IEEE 16th intl conf on dependable autonomic and secure computing 16th intl conf on pervasive intelligence and computing 4th intl conf on big data intelligence and computing and cyber science and technology congress (DASC/PiCom/DataCom/CyberSciTech). IEEE, pp 187–192
67. Pan JZ, Pavlova S, Li C, Li N, Li Y, Liu J (2018) Content based fake news detection using knowledge graphs. In: International semantic web conference. Springer, pp 669–683
68. Della Vedova ML, Tacchini E, Moret S, Ballarin G, DiPierro M, de Alfaro L (2018) Automatic online fake news detection combining content and social signals. In: 2018 22nd conference of open innovations association (FRUCT). IEEE, pp 272–279
69. Atodiresei C-S, Tănăsescu A, Iftene A (2018) Identifying fake news and fake users on twitter. *Procedia Comput Sci* 126:451–461
70. Wang Y, Ma F, Jin Z, Yuan Y, Xun G, Jha K, Su L, Gao J (2018) 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

71. Ajao O, Bhowmik D, Zargari S (2018) 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
72. Karimi H, Roy P, Saba-Sadiya S, Tang J (2018) Multi-source multi-class fake news detection. In: Proceedings of the 27th international conference on computational linguistics, pp 1546–1557
73. Wu L, Liu H (2018) Tracing fake-news footprints: characterizing social media messages by how they propagate. In: Proceedings of the eleventh ACM international conference on web search and data mining, pp 637–645
74. Gupta S, Thirukovalluru R, Sinha M, Mannarswamy S (2018) Cimtddetect: a community infused matrix-tensor coupled factorization based method for fake news detection. In: 2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM). IEEE, pp 278–281
75. Dong M, Yao L, Wang X, Benatallah B, Sheng QZ, Huang H (2018) Dual: A deep unified attention model with latent relation representations for fake news detection. In: International conference on web information systems engineering. Springer, pp 199–209
76. Guacho GB, Abdali S, Shah N, Papalexakis EE (2018) Semi-supervised content-based detection of misinformation via tensor embeddings. In: 2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM). IEEE, pp 322–325
77. Helmstetter S, Paulheim H (2018) Weakly supervised learning for fake news detection on twitter. In: 2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM). IEEE, pp 274–277
78. Liu Y, Wu Y-F (2018) Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In: Proceedings of the AAAI conference on artificial intelligence, vol 32
79. Shu K, Zhou X, Wang S, Zafarani R, Liu H (2019) The role of user profiles for fake news detection. In: Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining, pp 436–439
80. Reis JC, Correia A, Murai F, Veloso A, Benevenuto F (2019) Supervised learning for fake news detection. *IEEE Intell Syst* 34(2):76–81
81. Olivieri A, Shabani S, Sokhn M, Cudré-Mauroux P (2019) Creating task-generic features for fake news detection. In: Proceedings of the 52nd Hawaii international conference on system sciences
82. Bharadwaj P, Shao Z (2019) Fake news detection with semantic features and text mining. *Int J Nat Lang Comput* 8
83. Rasool T, Butt WH, Shaikat A, Akram MU (2019) Multi-label fake news detection using multi-layered supervised learning. In: Proceedings of the 2019 11th international conference on computer and automation engineering, pp 73–77
84. Traylor T, Straub J, Snell N et al (2019) Classifying fake news articles using natural language processing to identify in-article attribution as a supervised learning estimator. In: 2019 IEEE 13th international conference on semantic computing (ICSC). IEEE, pp 445–449
85. Shu K, Mahudeswaran D, Liu H (2019) Fakenewstracker: a tool for fake news collection detection and visualization. *Comput Math Organ Theory* 25(1):60–71
86. Yang S, Shu K, Wang S, Gu R, Wu F, Liu H (2019) Unsupervised fake news detection on social media: a generative approach. In: Proceedings of the AAAI conference on artificial intelligence, vol 33, pp 5644–5651
87. Kesarwani A, Chauhan SS, Nair AR (2020) Fake news detection on social media using k-nearest neighbor classifier. In: 2020 International conference on advances in computing and communication engineering (ICACCE). IEEE, pp 1–4
88. Shu K, Mahudeswaran D, Wang S, Liu H (2020) Hierarchical propagation networks for fake news detection: investigation and exploitation. In: Proceedings of the international AAAI conference on web and social media, vol 14, pp 626–637
89. Zhou X, Jain A, Phoha VV, Zafarani R (2020) Fake news early detection: a theory-driven model. *Digit Threat Res Pract* 1(2):1–25
90. Lu Y-J, Li C-T (2020) Gcan: graph-aware co-attention networks for explainable fake news detection on social media. In: Proceedings of the 58th annual meeting of the association for computational linguistics, pp 505–514. <https://doi.org/10.18653/v1/2020.acl-main.48>
91. Ni S, Li J, Kao H-Y (2021) Mvan: multi-view attention networks for fake news detection on social media. *IEEE Access* 9:106907–106917
92. Sahoo SR, Gupta BB (2021) Multiple features based approach for automatic fake news detection on social networks using deep learning. *Appl Soft Comput* 100:106983
93. Kaliyar RK, Goswami A, Narang P (2021) Fakebert: fake news detection in social media with a bert-based deep learning approach. *Multimed Tools Appl* 80(8):11765–11788

94. Dou Y, Shu K, Xia C, Yu PS, Sun L (2021) User preference-aware fake news detection. In: Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval, pp 2051–2055
95. Verma PK, Agrawal P, Amorim I, Prodan R (2021) Welfake: word embedding over linguistic features for fake news detection. *IEEE Trans Comput Soc Syst* 8(4):881–893
96. Ozbay FA, Alatas B (2021) Adaptive salp swarm optimization algorithms with inertia weights for novel fake news detection model in online social media. *Multimed Tools Appl* 80(26–27):34333–34357
97. Nasir JA, Khan OS, Varlamis I (2021) Fake news detection: a hybrid cnn-rnn based deep learning approach. *Int J Inf Manag Data Insights* 1(1):100007
98. Iwendi C, Mohan S, Ibeke E, Ahmadian A, Ciano T et al (2022) Covid-19 fake news sentiment analysis. *Comput Electr Eng* 101:107967
99. Seddari N, Derhab A, Belaoued M, Halboob W, Al-Muhtadi J, Bouras A (2022) A hybrid linguistic and knowledge-based analysis approach for fake news detection on social media. *IEEE Access*
100. Min E, Rong Y, Bian Y, Xu T, Zhao P, Huang J, Ananiadou S (2022) Divide-and-conquer: post-user interaction network for fake news detection on social media. In: Proceedings of the acm web conference 2022, pp 1148–1158
101. Galli A, Masciari E, Moscato V, Sperlì G (2022) A comprehensive benchmark for fake news detection. *J Intell Inf Syst* 59(1):237–261
102. Sadeghi F, Bidgoly AJ, Amirkhani H (2022) Fake news detection on social media using a natural language inference approach. *Multimed Tools Appl* 81(23):33801–33821
103. Chi H, Liao B (2022) A quantitative argumentation-based automated explainable decision system for fake news detection on social media. *Knowl-Based Syst* 242:108378
104. Choudhury D, Acharjee T (2022) A novel approach to fake news detection in social networks using genetic algorithm applying machine learning classifiers. *Multimed Tools Appl* 1–17
105. Shao C, Hui P-M, Wang L, Jiang X, Flammini A, Menczer F, Ciampaglia GL (2018) Anatomy of an online misinformation network. *Plos One* 13(4):0196087
106. Pérez-Rosas V, Kleinberg B, Lefevre A, Mihalcea R (2018) Automatic detection of fake news. In: Proceedings of the 27th international conference on computational linguistics. Association for Computational Linguistics, pp 3391–3401. <https://aclanthology.org/C18-1287>
107. Singh MK, Ahmed J, Raghuvanshi KK, Alam MA (2023) Bharatfakenewskosh: a data repository for fake news research in India. In: International conference on smart computing and communication. Springer, pp 277–288
108. Al-Rubaie A, Al-Jadir L, Al-Hashimi M (2019) Detecting arabic fake news using machine learning techniques. In: Proceedings of the 2019 3rd international conference on computer science and artificial intelligence (CSAI), pp 201–206
109. Beauvais C (2022) Fake news: why do we believe it. *Joint Bone Spine* 89(4):105371
110. Djenouri Y, Belhadi A, Srivastava G, Lin JC-W (2023) Advanced pattern-mining system for fake news analysis. *IEEE Trans Comput Soc Syst*
111. Raponi S, Khalifa Z, Oligeri G, Di Pietro R (2022) Fake news propagation: a review of epidemic models datasets and insights. *ACM Trans Web* 16(3):1–34
112. Ahmed U, Lin JC-W, Diaz VG (2022) Automatically temporal labeled data generation using positional lexicon expansion for focus time estimation of news articles. *ACM Trans Asian Low-Resour Lang Inf Process*
113. Ahmed U, Lin JC-W, Srivastava G, Yun U (2023) Semi-supervised lexicon-aware embedding for news article time estimation. *ACM Trans Asian Low-Resour Lang Inf Process*
114. Raza S, Ding C (2022) Fake news detection based on news content and social contexts: a transformer-based approach. *Int J Data Sci Anal* 13(4):335–362

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Manish Kumar Singh¹  · Jawed Ahmed¹ · Mohammad Afshar Alam¹ · Kamlesh Kumar Raghuvanshi² · Sachin Kumar³

Jawed Ahmed
jahmed2047@jamiyahamdard.ac.in

Mohammad Afshar Alam
aalam@jamiyahamdard.ac.in

Kamlesh Kumar Raghuvanshi
kamlesh@ramanujan.du.ac.in

Sachin Kumar
skumar@cic.du.ac.in

¹ Department of Computer Science Engineering, School of Engineering Science & Technology, Jamia Hamdard University, New Delhi 110062, India

² Department of Computer Science, Ramanujan College, University of Delhi, Kalkaji, New Delhi 110019, Delhi, India

³ Cluster Innovation Center, University of Delhi, GC Narang Road, New Delhi 110007, Delhi, India