Knowledge and Information Systems (2023) 65:3839–3880
https://doi.org/10.1007/s10115-023-01902-w

REVIEW

A survey on rumor detection and prevention in social media using deep learning

Barsha Pattanaik[1] · Sourav Mandal[1] · Rudra M. Tripathy[1]

Received: 5 December 2022 / Revised: 29 April 2023 / Accepted: 9 May 2023 /
Published online: 29 May 2023

Abstract

In the current digital era, massive amounts of unreliable, purposefully misleading material, such as texts and images, are being shared widely on various web platforms to deceive the reader. Most of us use social media sites to exchange or obtain information. This opens a lot of space for false information, like fake news, rumors, etc., to spread that could harm a society's social fabric, a person's reputation, or the legitimacy of a whole country. Therefore, preventing the transmission of such dangerous material across platforms is a digital priority. However, the main goal of this survey paper is to thoroughly examine several current state-of-the-art research works on rumor control (detection and prevention) that use deep learning-based techniques and to identify major distinctions between these research efforts. The comparison results are intended to identify research gaps and challenges for rumor detection, tracking, and combating. This survey of the literature makes a significant contribution by highlighting several cutting-edge deep learning-based models for rumor detection in social media and critically evaluating their effectiveness on recently available standard datasets. Furthermore, to have a thorough grasp of rumor prevention to spread, we also looked into various pertinent approaches, including rumor veracity classification, stance classification, tracking, and combating. We also have created a summary of recent datasets with all the necessary information and analysis. Finally, as part of this survey, we have identified some of the potential research gaps and challenges that need to be addressed in order to develop early, effective methods of rumor control.

Keywords Rumor detection · Rumor veracity classification · Rumor tracking · Rumor combating · Deep learning

✉ Barsha Pattanaik
barsha@xustudent.edu.in

Sourav Mandal
sourav.mandal@ieee.org

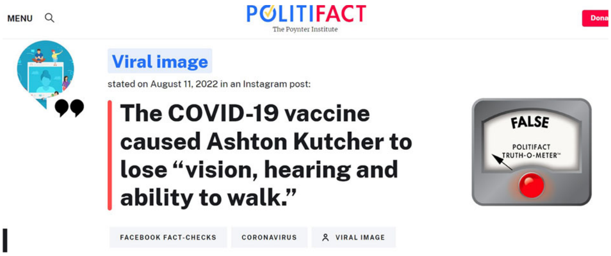
Rudra M. Tripathy
rudramohan@xim.edu.in

[1] School of Computer Science and Engineering, XIM University, Bhubaneswar, Odisha 752050, India

1 Introduction

The internet has permeated every aspect of our lives in the current era. To share or subscribe to information for commerce, health care, education, and politics, people use a variety of social networking sites. Huge amounts of information are being produced and disseminated throughout social networking platforms as a result. The proliferation of information can sometimes have a negative impact on society by upsetting the social order and making people more confused. Social media is being used to post a lot of stuff. Not all information about the occasion or organization is accurate; some of it could be factual, false, or unreliable [28]. People do not hesitate to propagate their destructive misinformation by abusing social media platforms. Without validating the accuracy of the material, social media outlets, including Facebook, Twitter, Instagram, television channels, emails, have been used to cascade the spread of information [49]. This promotes the dissemination of erroneous, fraudulent, or questionable information [12, 13]. Due to this, enormous amounts of information are produced and disseminated on social networking sites. Sometimes, information overload can be bad for society because it throws off the social order and confuses people. Such information fraud negatively affects people, groups, and countries, either directly or indirectly. This type of false information or news is referred to as "misinformation" because its primary goal is to deceive the common public. Unreliable, unsubstantiated, and unattributed rumors, which is a type of misinformation, pose a major hazard because they can be either real or untrue. Because rumors are so deceiving, there is a greater need for rumor detection models so that they can stop rumors from propagating through social media and news websites. In general, misinformation causes misunderstandings among people, which have an impact on practically all fields [24]. Although it is not a brand-new issue, digital media has made information sharing quick and simple, which makes it harder to access the real information. The spread of false information frequently results in misunderstandings, which negatively affects someone's life and occasionally results in suicide [45]. It can appear in a variety of ways, such as spam, false information, deception, and fake news [35]. The main distinction between fake news, false news, disinformation, rumor, and misinformation is the reason for its dissemination [101]. All four terms are practically the same and interchangeable. In fact, it can be difficult to determine the veracity of each piece of information [49]. The key differences among them are the creators' intentions, the impact on society, and the spread's media, direction, and speed. Since social media allows for the real-time sharing of information and news, the impact of any malicious behavior, such as the dissemination of misleading information or unconfirmed photographs, must be recognized and curbed. Such erroneous information spreads throughout society, causing instability and anxiety. There are several websites, but two well-known ones are www.snopes.com and www.politifact.com, where all kinds of recent news, stories, and events are classified as rumors or not. These websites have categorized the events as true or false after looking into the facts. According to Fig. 1, Ashton Kutcher allegedly lost his ability to hear, see, and walk after receiving the COVID-19 vaccine. But a few days later, the actor made the official announcement that, prior to the release of the COVID-19 vaccination, he had been diagnosed with vasculitis. After looking at these issues, the PolitiFact website declared this rumor to be false. These two websites identify two other posts as non-rumor and rumor, shown in Figs. 2 and 3, respectively.

In this survey, we have reviewed recent literature with the aim of thoroughly comprehending the rumor detection and control approaches. We have observed that while there has been some work done and published on rumor detection, not as much has been done on fake news detection. We have also observed a dearth of useful survey papers on rumor that

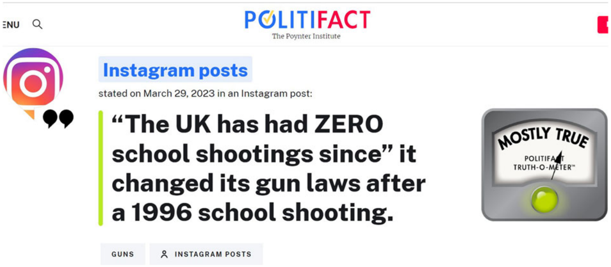


Viral image

stated on August 11, 2022 in an Instagram post:

The COVID-19 vaccine caused Ashton Kutcher to lose "vision, hearing and ability to walk."

Fig. 1 Marked as rumor from www.politifact.com

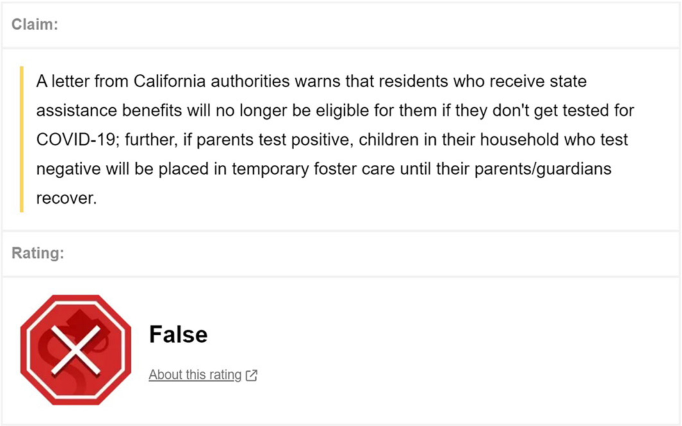


Instagram posts

stated on March 29, 2023 in an Instagram post:

"The UK has had ZERO school shootings since" it changed its gun laws after a 1996 school shooting.

Fig. 2 Marked as non-rumor from www.politifact.com



Claim:

A letter from California authorities warns that residents who receive state assistance benefits will no longer be eligible for them if they don't get tested for COVID-19; further, if parents test positive, children in their household who test negative will be placed in temporary foster care until their parents/guardians recover.

Rating:

False

About this rating

Fig. 3 Marked as rumor from www.snopes.com

directly addressed DL-based methodologies while simultaneously outlining the challenges and research gaps in the literature. This encourages us to conduct a survey on rumor detection and control methods. This survey paper describes the varieties of DL-based models that modern rumor detection systems employ. According to state-of-the-art systems, there will be a collection of messages (or posts) from which the system needs to identify whether a particular message or a post is a rumor or not. For rumor veracity classification task, the system has to classify a message into any one of the four classes—true rumor, false rumor, non-rumor, and unverified (see Sect. 2.5 for details). However, Fig. 4 gives a pictorial representation of the generic rumor detection process pipeline (as binary classification). The first phase is to collect the data and to prepare the dataset. Most of the researchers use social

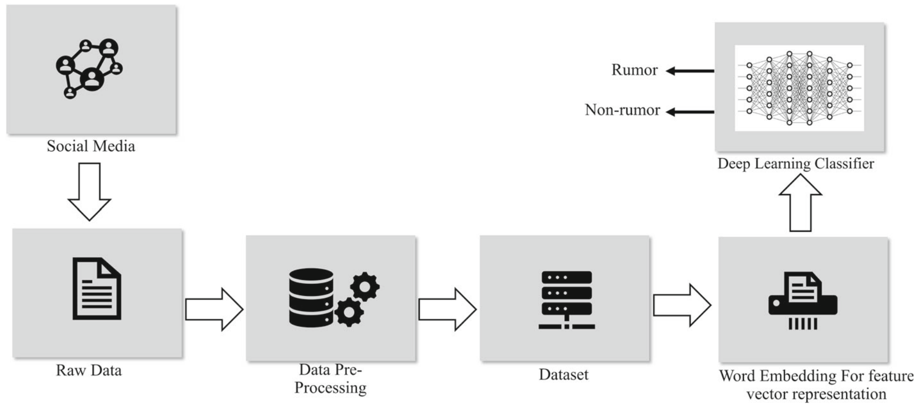


Fig. 4 Generic rumor detection process pipeline

media platforms or fact-checking sites to collect the data (posts or messages) and apply natural language processing (NLP)-based preprocessing techniques to prepare the input datasets. Some researchers directly used standard datasets for their experiments. After data collection, various word embedding techniques are applied for feature vector generation, and those feature vectors are fed to the DL-based models for training and for the final prediction, i.e., an input message is a rumor or a non-rumor (see Fig. 4).

Rumor detection is an intelligent task since it automatically assesses web material using rumor detection frameworks, tools, and plugins. By showcasing several cutting-edge DL-based models for rumor detection in social media and critically assessing their performance on recently made standard datasets, this literature survey significantly contributes to the field. We have gathered research articles from reputed journals and conferences, which are listed in the bibliography, in order to understand the current research state of the rumor detection models and control mechanisms. In addition, this analysis makes an effort to pinpoint some of the potential research gaps and challenges in rumor detection, stance and veracity classification, tracking, and combating that must be overcome in order to create prompt, efficient rumor control strategies. In the next subsection, we discuss a few of the existing survey papers on rumor detection to comprehend their contributions and contents better. A number of our most significant contributions to this entire survey paper are then listed in Sect. 1.2.

1.1 Existing survey papers on rumor detection techniques

Researchers have created a number of models using artificial intelligence (AI), including natural language processing (NLP), machine learning (ML), and deep learning (DL) algorithms, to automatically recognize various types of misinformation such as fake news and rumors. There have been numerous studies on various misinformation types. As a result, it is challenging for new researchers to determine the current status of this research. In order to assess the current status of this research in spotting these numerous claims, detailed literature survey is recommended. We discovered a few survey papers that had previously been written on the same subject [3, 12, 35, 87, 91, 118]. We looked at them to understand the state of the research at the time and discovered that some important information related to some recently published articles and evidence of research gaps were missing. This motivated us to thoroughly examine this subject using relevant published works, and we ultimately critically analyzed some potential research efforts to pinpoint research gaps.

For instance, Bondielli and Marcelloni [12] surveyed many research works on detection of fake news and rumors both. Along with the datasets utilized by other studies, the authors described various features and techniques used to detect fake news and rumors. Similar to this, Islam et al. [35] did a thorough survey on various misinformation detection methodologies based on different DL-based techniques in order to outline the future research route for new researchers. Tan et al. [87] offered a thorough, in-depth assessment on rumor detection using DL techniques. The authors outlined the DL-based rumor detection models and categorized them into different groups. They also explained how to select different features from the datasets and use them for rumor detection. Finally, the authors explained different datasets along with some future research directions on rumor detection. However, the authors ignored other components to control rumors in social media such as how to track and combat rumor. Zubiaga et al. [118] did a survey on social media rumors where they mentioned the whole rumor classification system starting from rumor detection, rumor tracking, stance, and veracity classification. Along with this, the author explained various types of rumors from different perspectives, along with some future research directions for rumor detection. However, the author did not mention any methods or techniques to prevent rumors on social media. Al-Sarem et al. [3] did a systematic literature review on rumor detection to provide a deep understanding of how researchers used different DL methods to detect rumors in social media. Along with this, the authors also mentioned the challenges the researchers face while working in this area. Finally, the authors provided some prospective future research directions for rumor detection. However, they ignored other aspects in rumor control like rumor tracking and combating in social media. In another survey, Varshney and Vishwakarma [91] provided a detailed survey on rumors by focusing few aspects, such as how to access data from different social media, and the features used by researchers in different models. The authors also surveyed different methods for rumor detection and veracity classification from text and image-based rumor data. In their survey, they highlighted some intriguing research studies on rumor detection from images. Our survey paper represents the effort to analyze and explain similarities and differences between various types of rumor detection models that used DL-based approaches and many more. It contributes insights from a variety of angles, some of which are described in the following subsection.

1.2 Significant contributions of this survey

We have paid special attention to rumors and the methods used to detect them, including deep learning. We have also talked about tracking rumors, classifying stance (or positions), classifying veracity (or authenticity), and combating rumors, among other topics. However, the primary contributions of this survey paper that set it apart from other surveys are outlined below.

- We have done a detailed and complete literature review up to date on the topic of rumor detection and its prevention that used DL-based approaches. We have grouped the recent articles according to various DL-based methodologies, such as recurrent neural network (RNN), convolutional neural network (CNN), graph-based convolutional network (GCN), and hybrid models, as well as the most recent state-of-the-art techniques (see Sect. 2).
- Misinformation such as fake news and rumor, its effects on society, and its associated terminologies have all been examined. Additionally, we have discussed the differences

and similarities between rumor and fake news as well as how they are both subsets of misinformation. We have included a brief assessment of the research on fake news detection using several DL-based techniques as these techniques are quite similar to detect rumors as well (see Sect. 1.6).

- We have also examined all the datasets currently available for rumor detection and summarized how the researchers' proposed models performed on those datasets (see Sects. 3.1 and 3.3).
- We have also included a brief survey on rumor stance classification, veracity classification, rumor tracking, rumor combating as these are also interesting research areas to work along with rumor detection. Other similar survey papers lack such depth of detailing (see Sect. 2).
- This paper concludes with a summary that includes a critical analysis of the potential research works and any remaining research gaps and challenges (see Sect. 3.3). None of the surveys on rumor like [35, 87] have specifically pointed out the research gaps and challenges that are crucial for the development of any research in the future.

In the following subsection, we will discuss the concept of misinformation and a technical explanation of what fake news and rumor are.

1.3 Misinformation

Misinformation is a false information which is intentionally spread to deceive people [35, 45, 69]. Zhou et al. [113] explained that information which is inaccurate, unsupported by facts, and intended to mislead the public is also referred to as misinformation. When people spread a steady stream of erroneous and misleading information, misinformation is generated [85]. Misinformation is also produced when people decide to omit true information and spread falsehoods [35]. It is also known as falsehood, obscurity, deception, etc. [78]. Misinformation is sometimes referred to by phrases like fake news, rumors, disinformation, and spam [35]. Fake news, rumors, disinformation, and spam are all types of misinformation, according to Wu et al. [98]. In this work, we have mostly concentrated on rumors, fake news, and misinformation overall, which are all related to one another as illustrated in Fig. 5.

Fake news News which is manufactured and presented as real fact in an effort to deceive the reader is referred to as fake news [13]. For instance, in 2016 "Buzz Feed News" classified a news—"Barack Obama signed an executive order to ban the pledge of allegiance in schools nationwide[1]" as fake news.

Rumor Rumors are described as unverified, doubtful information that has been spread by someone [12]. For instance, there was a report on Twitter a few months ago that "#100days100nights" according to it, a contest to see "who would be the first to kill 100 people" was held in the Los Angeles area. The population of south Los Angeles was so quickly exposed to this message that no one dared to venture outside for fear of becoming a victim. [2]

1.4 Fake news and rumor—a subset of misinformation

Islam et al. [35] categorized misinformation into many categories. Fake news has been classified by the authors [35] as a subset of misinformation. A modified version of an original news

[1] https://www.snopes.com/fact-check/pledge-of-allegiance-ban/.

[2] https://www.snopes.com/fact-check/gangs-kill-100-days/.

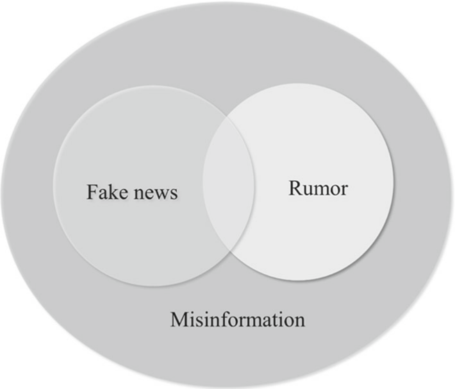


Fig. 5 Fake news and rumor—a subset of Misinformation

report is all that fake news is, and it is quite tough to spot. In a similar vein, Bondielli and Marcelloni [12]'s review paper explains the categorization of misinformation and demonstrates the fake news and rumors. Wu et al. [98] gave a thorough breakdown of many sorts of misinformation. The author claims that the term "misinformation" serves as a catch-all for all forms of incorrect information. Since many social media users fail to verify any misleading information and unintentionally distribute it. Following [98], Islam et al. [35] classified false information into a number of terms, which include fake news and rumors. Bondielli and Marcelloni [12] divided false information or misinformation into three categories, including rumors, fake news, and some other types related to misinformation on social media. The author defined misinformation as an inaccurate information that is disseminated online and through social media. The expressions "fake news" and "rumors" are examples of misinformation. Rumors are inaccurate information that has not been verified, whereas fake news is a serious fabrication. After reading the literature [12, 35, 98], we discovered that misinformation includes fake news and rumors as a subset. These two stories are both untrue. The deliberate dissemination of false information known as fake news takes the shape of real news. Rumors, on the other hand, are often unverified misinformation that occasionally may or may not be true. The influence of rumor on society in terms of panic is stronger, and it spreads quickly.

1.5 Similarity and dissimilarity between rumor and fake news

False, unverified, and questionable information is what meant by rumors [12], whereas fake news is described as created information that is presented as the truth [13]. Rumors and fake news are both untrue. Both rumors and fake news have the potential to harm society severely over very short periods of time. While rumors are unconfirmed information with no assurance, fake news is described as news that is deceptive and contains material that is either entirely or partially incorrect. These two forms of misinformation, however, share many characteristics (or features). The most striking connection between rumors and misleading information is that both have disastrous effects on society. If we compare fake news to rumor, we can observe that fake news is always released on purpose and is detrimental to society. Furthermore, when a rumor starts to spread, the general public is unaware of its veracity. Rumor has a considerably larger social influence than fake news. The hardest part is spotting erroneous information before it negatively affects society. Based on a few criteria, we contrast and compare the two. In order to assist academics researching on rumor detection, we have reviewed some of the most recent papers on detecting fake news which are also included in this paper (see Sect. 1.6). Three aspects, such as feature extraction, datasets used, and detection methodologies, are used to explain the similarities and differences between fake news and rumors.

1.5.1 Feature extraction

In terms of feature extraction techniques, fake news detection uses content-based techniques. Usually, the contents just copy text features [12]. Natural language processing (NLP) plays a vital role in extracting the essential information from the text in a systematic manner by analyzing semantic, lexical, and syntactic characteristics. Semantic feature analysis has typically been applied with ML and DL through NLP techniques to detect false information, such as fake news and rumors. Lexical feature analysis focuses on the words that are present in the text, whereas syntactic feature analysis focuses on the structure and complexity of the sentences. The method that is typically used to extract context from text is context-based. This kind of method examines the audience, the news's source, and the networks that allow for the spread of information, etc. NLP algorithms were used to directly extract information from the text in the instance of fake news, based mostly on the contents. Madani et al. [65] employed a new algorithm for classifying fake news based on COVID-19 news. The number of tweets, retweets, the source of the tweet, its length, the user, followers, and attitudes were all used by the authors as features in this experiment. In another paper, Huang and Chen [34] employed bi-grams and uni-grams as features in addition to do text and grammar analysis. But to extract relevant information from the real-world data, few researchers used context-based features for fake news detection. For example, to detect fake news in social media Shim et al. [80] used a context-based technique to extract features while analyzing network and user-based data.

However, rumor detection analysis can be used to extract features based on both content and context [12, 32]. Bai et al. [9] created a graph convolutional network (GCN) for rumor detection in social media by utilizing both local and global structural features between tweets and related replies. Local features are related to the relationship between the current tweet and the reply tweet with the source tweet, while global structural features referred to the relationship between all tweets in a conversation. In order to create a source and reply graph, which was employed for rumor detection analysis, the researchers used both of these structures as features.

1.5.2 Datasets

Naturally, gathering datasets for rumors and fake news is a difficult undertaking. Most of the researchers used publicly available benchmark datasets for their research work. Benchmark datasets are also very hardly available. The most popular publicly accessible fake news datasets are ISOT [2]. FakeNewsNet,[3] LIAR dataset [95]. PHEME [114], TWITTER [61, 62], and Weibo [61] are a few often available datasets that researchers have been using in the case of rumors. Collecting real-world datasets is a challenging task. Facebook, Twitter, and WhatsApp are the three most widely used social networking platforms. There are numerous ways to obtain the data on these platforms. The APIs of various social media platforms like Twitter have been used by many researchers to collect data. Some researchers also use other

[3] https://github.com/KaiDMML/FakeNewsNet.

debunking websites like Snopes, PolitiFact, Fact Check, etc. Few researchers also use Python libraries and Selenium web drivers to collect data from diverse websites [91].

1.5.3 Detection approaches

To identify fake news and rumors, DL and/or ML techniques are both used by the researchers. Fake news and rumor detection in social media primarily use ML techniques like regression, decision trees, support vector machines (SVM), naive Bayes, random forests, etc. Numerous researchers discovered that methods of ML produce accurate results in the classification of fake news. On the other hand, for greater performance, several researchers also applied a variety of DL techniques such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs). For example, Al-Sarem et al. [4] used a hybrid model consisting both CNN and RNN for rumor detection in social media. Similarly, Nasir et al. [68] also used a hybrid model containing both CNN and RNN for fake news detection. We found in our survey that most of the researchers used ML or DL techniques to detect fake news and rumor as well in social media. A thorough analysis of fake news and a survey on rumor detection is provided in Sects. 1.6 and 2.

1.6 A brief overview on fake news detection

Fake news is nothing new, and it is produced when people deliberately broadcast incorrect information to deceive society, which poses a greater threat to the country [45]. Any news or information that is entirely or partially false is defined as fake news [45]. Along with a survey on rumors, this paper also includes a comprehensive analysis of fake news to see how the two relate in terms of definition, dataset accessibility, identifying traits or features, and detection methods.

In order to put fake news detection strategies into practice, we read various articles on the topic. In a recent work, de Beer and Matthee [10] described many DL-based methods for spotting fake news on online social media. Further de Oliveira et al. [69] conducted a survey to find various approaches to preprocess the input data by using NLP technique. The author discussed about various methods of extracted data, by utilizing different ML algorithms. The authors also talked about potential areas for future investigation. Jwa et al. [40] proposed a DL-based model bidirectional encoder representations from transformers (BERT) for fake news detection by taking account the relationship between the body text and the headline of the news articles. The BERT was initially developed by Devlin et al. [22]. Pre-training language representations from unlabeled data were the focus of the model's [40] effort. The two procedures such as fine tuning of input message and pre-trained model were used to explain the proposed model. The initial phase was classifying the datasets into four groups—agree, disagree, discuss, and not related—using weighted cross-entropy. FNC-1[4] datasets were used for fine-tuning. Second, CNN and Daily Mails,[5] datasets were used to evaluate the model. Yang et al. [104] suggested a DL-based model to analyze both image and text data. In order to extract latent features from text and image input, the model consists of two CNN networks connected in parallel. Again to handle both textual and visual data, the authors established two branches. For the purpose of final classification, latent features that were derived from the CNN are used. The authors used the real-world dataset downloaded from

[4] https://github.com/FakeNewsChallenge/fnc-1.

[5] https://github.com/abisee/cnn-dailymail.

Kaggel.[6] The model gave 92.2% precision, 92.7% recall, and 92.1% F1 value to see details about performance evaluation matrices such as precision, recall, and F1 value (see Sect. 3.2).

Wu et al. [99] developed an adversarial network-based model to remove common and irrelevant aspects from the extracted features. Using BiLSTM and reinforcement learning, the most significant and constant words were collected from the input as the common features. The performance of the model was tested using the LIAR, Weibo, Twitter 16 datasets in this study, and text features as well as meta data aspects including subject, job, party, and speaker were also identified as crucial credibility factors. For the purpose of detecting fake news, Huang and Chen [34] suggested an ensembling model that combines long short-term memory (LSTM) and CNN based on different DL techniques such as embedding LSTM, depth LSTM, Linguistic Inquiry and Word Count (LIWC CNN), and N-gram CNN. These four techniques were normally used for word embedding based on LSTM and CNN. The authors employed the Buzz feed corpus [73] datasets to assess the model's performance. To describe the interaction between text and image data, Zeng et al. [108] created a unique model to explain the semantic relationship between text and image data. The model had of three modules. The first module's output served as input for the second and third modules. To ensure the similarity of two words, the text data were first tokenized, and then, vectorization was carried out to turn word tokens into word vectors by the Word2Vec word embedding technique. Then, using the ImageNet dataset, the VGG-19 [83] network was trained, and this previously trained model was employed once more to extract picture features. In the second module, the authors explored the differences between the words included in text and visuals using attention mechanisms and images that were related to the text. In the third module, a multimodal feature was used to record the correlation between text and visual modes. To determine whether the news was fake or not, the output of modules two and three were merged and input into the fake news detector. The authors employed two datasets, including as Twitter [21] and Weibo [39] to assess the model's performance. They demonstrated that their model outperformed other state-of-the-art models with accuracy of 75.8% for the Twitter dataset and 83.4% for the Weibo dataset.

For detecting fake news, Kong et al. [45] trained various neural network models. Data exploration was initially carried out after dataset collection (downloaded datasets from the www.Kaggel.com), when datasets were examined to prevent data imbalance. After that, word vectors were produced from news contents using n-gram approach. All these word vectors were then aggregated and turned to a single matrix, which was then used to train various neural networks. For a vital classification of fake news, Madani et al. [65] proposed a model focused on numerous aspects of the tweets, such as the source of the tweets and retweets, the name of the user, friends/followers, feelings, etc., in addition to the text of the tweets itself. The classification of the tweet was then done using several ML and DL algorithms to determine whether it is fake or real. Nasir et al. [68] proposed a DL-based hybrid model which consists of both CNN and RNN. For extracting local feature, CNN was used. Then, the output of the CNN was fed to the RNN. Long-term dependencies of the local features such as the title, topic, date, location, and sources of the news articles from the input articles were learned using RNN-based LSTM to classify the tweet as fake or real. Aslam et al. [7] proposed an ensemble DL-based model for categorizing fake news. In accordance with the features of the datasets, the authors used two DL models. Bidirectional LSTM (BiLSTM) network was used to extract textual features, i.e., statements from the datasets, whereas a deep dense model was applied to extract other features such as the title, subject of the news datasets. Finally, the two models were then combined for the final classification. The authors

[6] https://www.kaggle.com/mrisdal/fake-news.

used LIAR [95] dataset to evaluate the performance of the model. The model gave an accuracy of 89.8%, recall of 91.6%, precision of 91.3%, and F-score of 91.4% in LIAR dataset. Liu et al. [55] developed a deep triple network (DTN) model that makes use of knowledge graphs to help identify fake news and provide triple-enhanced justifications. Data from the 2016 US Election and the Brexit were used to assess the model's performance. The model's accuracy score of 94.8% was high compared to state-of-the-art models.

Islam et al. [36] proposed an autonomous model called "Ternion" to identify fake news by determining the veracity of news. To confirm the veracity of the news, the author of this paper concentrated on stance detection, author credibility verification, and machine learning-based fake news classification. Monti et al. [67] proposed a geometric DL model that operates on graphically structured data. This model deals with diverse data, such as user demographic information, content, news transmission, and social network structure. Shim et al. [80] presented a DL-based model to automatically identify the links that contain fake news stories from the results of web searches. To capture the semantic and long distance dependencies with in the input text sentence, Kaliyar et al. [41] proposed a BERT-based model. The model included one embedded layer, two dense connected layers, five convolution network layers, max pooling layers, and other layers as well. To explain the relationships between sentences in fake news identification, Wang et al. [96] investigated a graph-based DL model in conjunction with self-attention mechanisms. There are three steps in this model. An LSTM network was utilized to encode the input sentence in the first stage. Second, the relationships between far-flung sentences were taken into account, and the subsequent sentences were arranged in chronological order. Third, the output of sentence representation was successively fed to the LSTM network, which was then used to represent documents. The output is then passed to a max-pooling layer followed by SoftMax layer for classification. Yang et al. [104] examined the text and picture data combined and created a model utilizing CNN for fake news and fake image identification. They also demonstrated that their model provided 92.0% of precision, 92.0% of recall, and 92.0% of F1 value.

The datasets like Weibo, ISOT, FakeNewsNet, and LIAR are frequently utilized for the detection of fake news. To assess the effectiveness of their methodology, the researchers used a variety of evaluation indicators. Some of them assessed the accuracy, precision, and recall values to evaluate the success of their models. According to the survey, Aslam et al. [7] produced an accuracy of 89.8% on the LIAR dataset, whereas Zeng et al. [108] produced accuracy of 75.0% and 83.4% on the Twitter and Weibo datasets, respectively. Nasir et al. [68] employed two datasets to test the accuracy of their model, and they obtained 99.0% accuracy for the ISOT dataset and 69.0% accuracy for the FAKES [78] datasets. Furthermore, it is crucial to see if the trained model can identify fake news in new datasets that have not been used before. A number of researchers, including [68], are also trying to resolve this problem. They demonstrated that their model could successfully extract features from new fake news (to classify) that was not included in the training set. Again, fake news is not only found on text data; fake images, emojis, and videos are also used by people. Recently, Deepfake technology has caused a lot of concern regarding the spread of misinformation like fake news and rumor. Deepfake AI algorithms are employed to produce convincing audio, video, and image forgeries. Additionally, they produce wholly unique content in which individuals are depicted doing or saying things that they did not actually do or say. Deepfakes' capacity to disseminate fake news and rumor that looks to come from trustworthy sources poses the biggest threat. These methods have been used by many artists to create a variety of audio films that show well-known world figures making claims that they would never make in real-life situations [26]. To build deepfakes for images, videos, and audios, Gaur et al. [25] describe a detailed application of DL-based approaches. Therefore, employing deep neural networks to detect fake news or rumor from these items is undoubtedly a difficult challenge. We will see in the following section (Sect. 2) how comparable DL-based techniques are also used for rumor identification.

2 Rumor detection and prevention methodologies

This section will cover some recent research works majorly on rumor detection, with veracity and stance classification, and tracking and combating rumors. We conducted a thorough review of the literature and a critical analysis of the methodologies and accessible standard datasets for this aim. We begin with explanations of various systems and methodologies and then talk about the research's challenges and limitations. Finally, we compare and contrast different methodologies, critically examine their performances, and explore evaluation methods. This survey's goal is to pinpoint research gaps that exist in these works so that academicians can work on finding the best ways to eliminate them. Any online information that has not been independently confirmed or that is untrue is referred to be rumor [81]. According to Yang et al. [103], a rumor is merely an unverified story or explanation of events that circulates among people in relation to an occasion or object in society. For our culture, this has significant ramifications [41]. In the past, there was a report that onion and salt prices had increased in Bangladesh. Citizens of Bangladesh hurried to get these necessities without first performing any investigation [35]. There have been situations where rumors have destroyed people's reputations or the interests of the country. Therefore, it is critical to put a stop to such rumors before they have a chance to spread further.

Mechanisms must be developed to stop rumors before they damage the social fabric. To safeguard society from the threat of rumors, rumor detection techniques are quite limited [44]. Many academics are working hard to develop models that can identify rumors quickly and accurately. Before a rumor is discovered, it is necessary to evaluate it at several phases [49, 53, 107]. We will examine the phases of handling rumor in more detail in the next subsections.

2.1 Phases of handling rumor toward stop spreading

We defined multiple phases (or stages) of rumor handling, starting from detection to cease spreading, after reviewing a large number of research publications. The list of different phases is provided below.

1. Rumor detection
2. Rumor tracking
3. Stance classification
4. Veracity classification
5. Rumor combating

When a rumor is spread in social media, there are mostly four phases through which rumors have been detected and classified [118]. The first phase, according to Zubiaga et al. [118], is to determine whether a piece of information or a tweet is a rumor in social media. A few rumors are occasionally confirmed as being real and are then labeled as true rumors, but frequently unconfirmed information is later proven to be untrue. These rumors are referred to as false rumors, and they have an adverse direct and indirect impact on society. The next phase is to track (or monitor) the rumor, and the third is to determine the receiver's response after a rumor has been tracked. Checking a tweet's validity, or whether it is true, fake, or unverified, is the final phase. In our survey, we too followed the same phases. Additionally, we go over another phase, known as rumor combating, which is about how to stop rumors from spreading on social media.

We discuss each of these phases and how the researchers approached them in the literature in the following subsections and how the researchers used the approaches they outlined to try to solve the problems they ran into. We also investigate many datasets used by the researchers and critically analyze their work in order to determine the current research gaps.

2.2 Rumor detection

Rumor detection is the process of figuring out if a piece of information is true or false. According to certain research, the task of determining whether a narrative or online post is a rumor or not (i.e., a factual tale or a news piece) is known as rumor detection, and the task of evaluating the veracity of a rumor (true, false, or unverified) is known as rumor verification [43, 115]. This subsection will cover a thorough study of rumor detection.

2.2.1 Rumor detection—concepts, methodologies, and datasets

This section provides a thorough analysis of rumor detection based on the datasets used, methodologies, tools, and approaches employed in the study. We also highlighted the researchers' recommendations for future research as well as the difficulties they encountered while conducting the studies. In the end, we concentrated on careful evaluation of each paper that was refereed for this survey. We are motivated to conduct additional research in this area as a result of [87]'s rigorous survey on rumor detection. Our study reviews recent, intensely focused research on a number of topics, including datasets, methodologies/algorithms, techniques, outputs or results, performance evaluations, future scopes and extensions, and challenges related to them. On the basis of numerous DL networks, various models have been developed by diverse researchers. Some examples are the recurrent neural network, the convolutional neural network, and the graph convolutional neural network. Since they do not support manual feature extraction processes, DL methods have comparative advantages over ML ones. These are explained in the following subsections.

2.2.2 Recurrent neural networks (RNNs)-based approaches

The RNN essentially functions as a feed-forward deep neural network that handles time series or sequential input [61]. They are commonly employed in NLP-based applications like language translation, speech recognition, image captioning, etc., to resolve ordinal or temporal problems. When identifying rumors, researchers deploy a variety of RNN versions. Below is a brief discussion about them.

1. Long short-term memory (LSTM): An example of an RNN is the LSTM, which is used to identify long-term dependencies in the input data and aids in the retention of previous information. A cell with an input gate, a forget gate, and an output gate make up an RNN unit [31]. The LSTM network's gates control the flow of data into and out of the system. LSTM networks are frequently employed in a variety of applications, including sentiment analysis, text categorization, speech recognition, etc.
2. Gated recurrent unit (GRU): A reset gate and an update gate make up the GRU. It was first introduced by Cho et al. [18]. GRU, unlike LSTM, does not have an output gate. Speech modeling and handwriting recognition are few of the applications with GRU.

3. Bidirectional LSTM (BiLSTM): A sort of RNN called BiLSTM [75] has two LSTM networks and allows information to flow in both forward and backward directions.

Naturally, it is difficult to automatically disprove rumors at the beginning. Detecting rumor in the very beginning stage is termed as early rumor detection [15]. While we are discussing about social media platforms, microblogs are the most important platform where people can easily spread false news. So, it is important to identify those false news before it puts bad impacts on society or individual people.

In order to detect rumors from microblogs, Ma et al. [61] devised a model based on RNN that learns the hidden representations that absorb contextual information of input data over time. Three different RNN types, including LSTM, GRU, and tanh RNN, were used by the authors. The main purpose of the tanh RNN was to record the context of the data across time. The authors constructed two datasets for this experiment, using Twitter and Sina Weibo websites [61]. The authors also demonstrated how their approach provided sufficient accuracy for spotting the rumor in its infancy. For the Twitter dataset, the GRU unit with two layers (GRU-2) performed better, with an accuracy rate of 88% and F1 values of 89.8% and 86.0% for rumor and non-rumor data, respectively. Similarly, GRU-2 provided nearly 91.0% accuracy, 87.6% precision, 86.0% recall, and an F1 value of 91.4% for Weibo datasets.

Similarly, Chen et al. [15] proposed a RNN-based model by learning latent representation from the series of input posts and used the datasets, Twitter and Weibo constructed by Ma et al. [61] to detect the rumor from a relevant post. The model first transforms each event's input post into a feature matrix. The network has then used an attention technique to learn the latent representation. Each input passed through the feed-forward RNN is weighted by the attention mechanism. In the output hidden layer, there was finally a sigmoid function to classify whether or not the event was a rumor. The model gave 74.02% and 71.73% of precision and 68.75% and 70.34% of recall for Twitter and Weibo datasets.

A deep neural network-based model that examined a multitask learning scheme with linked framework was proposed by Ma et al. [63]. In a novel way, the authors handled two tasks, such as stance classification (see Sect. 2.4) and rumor detection, simultaneously using an RNN. The problem had been delineated in terms of two RNN-based neural network models in this research. For the first model, a single shared hidden layer was used, whereas the second model made use of a single task-specific hidden layer. The hidden unit was modeled using a GRU network to improve efficiency. The Twitter dataset was used for rumor detection, and PHEME dataset was used for stance classification. The common and task-unvarying aspects of both tasks are extracted, while they are being trained simultaneously, but each task still has its own unique set of features. Li et al. [54] proposed a multitask learning system that is applied to rumor detection. A sharing layer and a dual task-related layer make up the RNN model. To identify rumors, the authors relied on a trustworthy user, official spokesperson, news source, or expert on the topic of the rumor. The essential tweets were given top priority using the attention mechanism-based LSTM. The proposed model has been tested successfully on two datasets from RumorEval [20] and PHEME [114, 117]. The authors showed that their model outperformed other state-of-the-art models with an accuracy of 63.8% in the RumorEval dataset and 48.3% in the PHEME dataset.

For the purpose of early rumor identification, Liu et al. [57] proposed a model that combined a CNN pooling layer and an LSTM. In this paper, three different factors were taken into account for the detection process, such as the content of the news that are transmitted, the people who disseminate the rumor (spreaders), and the structure of the rumor diffusion. Information about the spreaders' influence and popularity was recorded. The information tree's dynamic behavior was also recorded to represent the diffusion process. Cheng et al. [17]

put up a rumor classification model based on their numerous rumor detection, tracking, and verification tests. An LSTM-based variational autoencoder model was used to extract latent representations of tweet-level text. blueThe model was collaboratively trained to extract pertinent, information-rich, appropriate, and user-friendly latent representations for each component of the rumor categorization.

A model for rumor identification in social media was proposed by Sujana et al. [86] by merging the original tweet and the comments to create a complete event. The model is a hierarchical model made up of a multi-loss function BiLSTM network. The phrase "embedding" was employed in the first phase to create post- and event-embedded vectors. Effective multitask learning was achieved using the multi-loss function. Because of its use, model training was expedited, but accuracy suffered as a result. The authors claimed that their model could manage data with variable lengths. The model had a 92.0% accuracy for the PHEME dataset. The authors also proved that their model was also used for rumor detection at the very beginning stage. They also demonstrated that their model may be used to identify rumors before they spread.

Contemporarily, Chen et al. [16] suggested a Bi-GRU-based model for early rumor detection. The context relations and sequence relationships between the two postings can be discovered using Bi-GRU. In order to improve the accuracy of rumor detection, this paper explains the data preprocessing method based on account filtering and text standardization, where the account filtering method was used to remove the junk accounts and text standardization was used to standardize tweets in the Twitter standard dataset. To evaluate the performance of the model, the authors used Twitter dataset. The model gave 86.3% of accuracy, 85.8% of recall, and 83.6% F1 value for Twitter dataset.

Wang et al. [94] presented a methodology based on reinforcement learning to identify rumor in its infancy. The model is composed of a detection model and a control model. The proposed method, a Q-learning network built on an RNN, managed the control model. To capture prospective features and sequential properties of the information, the control model had a LSTM network. The model gave superior F1 value of 81%, recall of 79.0%, precision of 79.0%, and accuracy of 81.0% on PHEME datasets.

BiLSTM network was proposed by Cen and Li [14] for automatic rumor identification in social media. Initially, the input text data were used to extract semantic features by multiple BiLSTM network. Second, the model retrieved three different categories of social features, including user information, social features of communication material, and Weibo content-based features. For vectorization, two different word vectors, such as skip gram and continuous bag of words (CBOW), were used. The output layer includes a SoftMax layer for classification. The model's accuracy, recall, precision, and F1 value were all greater than the state-of-the-art models at 95.0%, 94.5%, 94.5%, and 94%, respectively. Figure 6 depicts the model's abstract view from [14].

Ahmad et al. [1] suggested social- and content-based rumor detection tools for social media platforms. In contrast to cutting-edge baseline traits, the proposed features, in the author's opinion, are more useful in classifying rumors. This methodology for rumor detection was straightforward but effective. Most early rumor detection research concentrated on persistent rumors and presupposes that rumors are always untrue. Using a bidirectional LSTM-RNN classifier, a DL model was applied to text data. In comparison with the state-of-the-art models, the proposed model produced the best results, with 87.0% precision, 88.0% recall, and 87.0% F1 score for PHEME datasets.

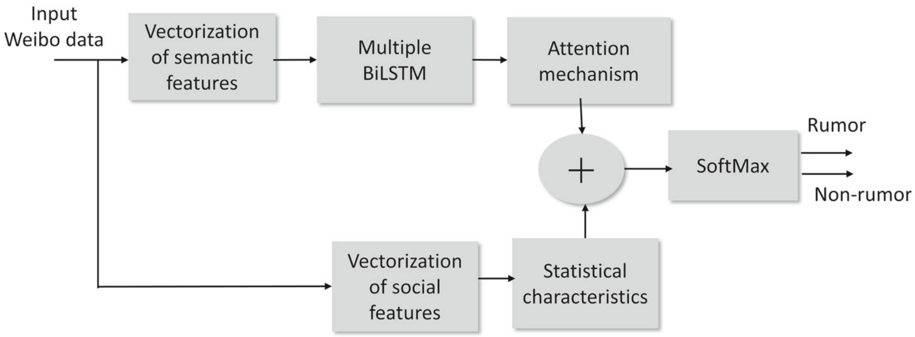

Input
Weibo data
Vectorization of semantic features
Multiple BiLSTM
Attention mechanism
+
SoftMax
Rumor
Non-rumor
Vectorization of social features
Statistical characteristics

Fig. 6 BiLSTM model

2.2.3 Convolutional neural network (CNN)-based approaches

Though RNN were mostly used to handle sequential text input, several researchers have lately started using CNN on text data alone or in combination with RNN. As an unsupervised feed-forward neural network with three layers-input, output, and several hidden layers is known as CNN [87]. Convolutional layers with nonlinear activation functions, pooling layers, and fully connected (dense) layers [5] are examples of hidden layers.

Alsaeedi and Al-Sarem [5] proposed a CNN-based rumor classification model. The authors used PHEME datasets for this purpose. Initially, the datasets were divided into training and testing set. The tokenized tweets were converted to vectors sequentially through word embedding. Then, the output of the embedding layer fed to the CNN model. The model had one convolutional layer with one pooling layer. The output layer had a "Sigmoid" activation function to predict whether the input tweet is rumor or non-rumor. The model gave an accuracy of 86% on PHEME datasets.

A deep transfer model based on the stochastic gradient descent algorithm was suggested by Guo et al. [27] to detect rumor in social media. In this paper, the model parameters, obtained during the training phase on the polarity review data, were given to the rumor detection model as input. The datasets used in this investigation were Five Breaking News (FBN) [116] and Yelp Polarity (YELP-2) [111]. The proposed model outperformed in terms of accuracy 87.28%, precision 79.12%, and F1 value 82.5%. In their system for rumor identification, Tu et al. [90] combined CNN with a propagation structure to incorporate features in the contents of text data of tweets and transmission topology of multiple tweets into a single graph. Two branches, such as the text branch and the node branch were used to make the model. They changed the text in the original tweet into a representation of word embedding for the text branch's input. The propagation sequence for the node branch's input was created by identifying the users that shared the tweet, then, this sequence was converted into the embedding representation of user nodes sequence. The altered propagation sequence and the source tweet's high order features were then extracted using CNN in both branches. The result from both branches was then combined to form a single vector. This vector was fed into two more layers-a fully connected layer and a SoftMax layer. The final output was represented in terms of the probability distribution over the set of classes for the associated tweet. Three datasets, including Twitter 15 [62], Twitter 16 [62], and Weibo [61], were used in the experiment. While the model's accuracy for the Weibo dataset was 95.0%, it was 79.0% and 85.0% for the Twitter 15 and Twitter 16 datasets, respectively, which was obviously greater than the accuracy of the models. See the Fig. 7 for detailed workflow.

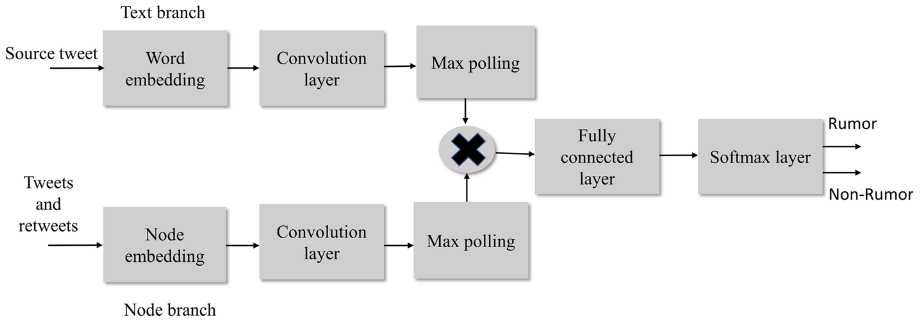


Fig. 7 Rumor2vec model

2.2.4 Graph convolution network (GCN)-based approaches

The social relationships between diverse items can be depicted in terms of data using the intricate structure of the social network. A type of neural network called a "graph convolution network" is used to extract the global features from a graph [87]. Normally, global features were extracted from the global structural information which is obtained by analyzing all the tweets with in a conversation. Compared to other DL networks such as RNN and CNN, GCN is more effective at extracting features of nodes in social networks [9].

A propagation tree was used as the input source for the two recursive neural network models that Ma et al. [64] proposed, with each node of the tree referring to the response post. The goal of this effort was to capture temporal aspects. The post's semantic features and their relationship to one another were extracted. The top-down model, which describes the information flow from the source post to the current node, and the bottom-up model, which describes the flow of information from the bottom-most leaves to the top-most source post, were both investigated in this study. The authors used two datasets Twitter 15 and Twitter 16. The accuracy was 72.3% for Twitter 15 and 73.7% for Twitter 16 datasets. For rumor identification via social media, Huang et al. [33] proposed a hybrid model that learns user representation through graph convolutional networks. User encoder, propagation tree encoder, and integrator are the three modules that make up their model. In order to learn a representation of the user, the user encoder uses a graph created by user behavioral data and GCN. The propagation tree encoder, which connects the content semantics with propagation hints, simultaneously learns the representation of the propagation tree using a recursive neural network. A fully connected layer with an integrator uses the output of the aforementioned module to detect rumors. Twitter 15, Twitter 16 datasets were used in this paper. The proposed model gave an accuracy of 75.2% for Twitter 15 and 77.3% for Twitter 16 datasets.

The primary characteristics of a rumor are transmission and dissipation of rumor [11]. To give a high level representation of both rumor transmission and dissipation, Bian et al. [11] proposed a graph-based neural network in two directions utilizing a top-down and bottom-up method. In this rumor classification model, a propagation structure based on the correlation between retweet and reaction was first built. Following that, features of propagation and dispersion were extracted using both top-down and bottom-up approaches. Then, the output of two approaches merged through a fully connected layer to achieve the final output. The classification of rumors was completed by depicting both the propagation and dissipation from the node structure. In order detect the rumors at its early stage, the authors focused on the structural features of both the propagation and dispersion of rumor. To evaluate the performance of the model they used three datasets such as Weibo, and other two Twitter 15 and Twitter 16 datasets. Finally, the author compared the performance value of the proposed model to other models including the model developed by Ma et al. [64] found that their model gave better accuracy, i.e., 96.1% for Weibo and 88.6% and 88.0% for Twitter 15 and Twitter 16 datasets. Similarly, Wei et al. [97] proposed edge enhanced Bayesian GCN model to analyze the content of the text data as well as their propagation structure in both the direction, i.e., both top-down and bottom-up approaches. Word embedding was used to embedding text data. Then both the node and graph modules were used to extract the structural features from the propagation graph of rumor. To evaluate the model, the authors used the same datasets used by [64] along with PHEME dataset where they found 89.2% of accuracy for Twitter 15 and 91.5% for Twitter 16 and 69.0% for PHEME datasets. The authors proved that their model was able to detect the rumors in its early stage.

A GCN-based ensembling model that takes into consideration the relationships between all tweets connected to a single topic was proposed by Bai et al. [9]. The original tweet is known as the "source tweet," and the proposed model is based on conversions between the original tweets and the replies. They had a graph that represented their relationship. The authors considered both the local and global structure of the input data. Conversations serve as the system's input, which is then converted into a word vector via a word2vec. The CNN and GCN were then given these word vectors. The author developed the model with text CNN layer and graph convolutional layers, a sort pooling layer and a Convolution layer. To learn the relation between source and reply tweets, graph spatial-based GCN was used and for text classification Text CNN was used. Finally, the output layer for classification is ensembled with the Text CNN and the spatial features-based GCN. For this experiment, the authors used PHEME datasets, which provide up-to-date information on five incidents. Overall output performance for the model was 85% of accuracy on PHEME dataset.

According to [84], most of the researchers while developing model for rumor detection only focused on content and propagation structure but were not aware of malicious attacks. It is a challenge to develop a model of adversarial attack by taking both the structural and textual information. According to them, two most important challenges were robustness to different responses and vulnerability to malicious attack. In order to overcome those challenges, Song et al. [84] proposed a GCN rumor detection model where first, a transfer-based encoder was utilized to encode each token in the whole discussion thread in order to leverage on the pre-trained information that was already present. Secondly, an adversarial response was added to the conversation thread. The model builds an adversarial response in the setting of a white-box attack, in which the detector's parameters and gradients are made public whenever the attacker is updated. PHEME, Twitter 15, and Twitter 16 datasets were used in this work. The model gave an accuracy of 93.47% for Twitter 15 and 90.19% for Twitter 16 and 84.84% for PHEME datasets.

Yang et al. [105] proposed the GCN model for social media rumor detection. The relationship between a shared post and the related posts were analyzed in this paper through a graph. The impact of any comments was then evaluated using a self-attention technique. In order to get the global representation, a post-related comments co-attention approach was introduced. We monitored the topical sway in the comments using a CNN. Topic drift is the evaluation of a topic that has been discussed over time to inform people of the reality of a situation. CNN's feed was utilized as a local representation. In this model, the rumor detection classification was carried out by concatenating the global and local representations of the data. To evaluate the performance of the model, three datasets were used and got the accuracy of 94.2% for Rumdect Ma et al. [61] and 95.4% for Weibo and 78.9% for Gossip Cop [82] datasets. Frequently, rumors spread from one place to another. Wang et al. [93] therefore

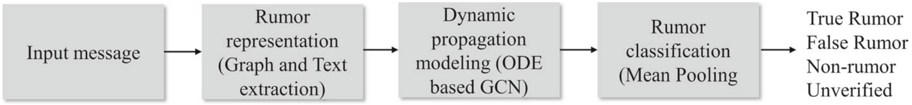

Input message
Rumor representation (Graph and Text extraction)
Dynamic propagation modeling (ODE based GCN)
Rumor classification (Mean Pooling
True Rumor
False Rumor
Non-rumor
Unverified

Fig. 8 Heterogeneous GCN model

proposed a GCN-based model to examine the features connected with the propagation zone in order to learn the pattern of rumor transmission. The authors used a source text-enhanced GCN to improve the model of the propagated message's learning capacity. In this study, the model takes the input data and extracts the textual features. For this research, two real-world datasets, including Twitter 15 and Twitter 16, were used. For Twitter 15 and 16, the model's accuracy was 85.6% and 87.8%, respectively.

To deal with the dynamic aspects, including the heterogeneous information, in rumor detection, Yu et al. [106] presented a GCN-based technique (see Fig. 8). The model is made up of three parts, including rumor classification, rumor transmission, and rumor representation. The rumor's content was first represented using TF-IDF, and the propagation node's vector was then encoded using an adjacency matrix. Then, a heterogeneous graph was created by combining the two. The GCN, which contains an ordinary differential equation (ODE), was then analyzed by this model. Finally, the classification process used hidden dynamic characteristics. The authors used Twitter 15 and Twitter 16 as two real-world datasets for this experiment. Compared to other models, The model gave an accuracy of 86.5%. Figure 8, which was adapted from the original graphic given in [106], provides a summary of the model.

2.2.5 Hybrid and ensemble model

Many researchers created hybrid or ensemble models by combining two or many strategies in order to improve the performance of the model during rumor detection. The term "hybrid model" refers to a model that incorporates two DL techniques, such as CNN or LSTM. In a hybrid model, the models often feed their output to one another to produce the result, whereas an ensemble model can predict an outcome on its own. A hybrid model based on DL was proposed by Kumar et al. [50] to develop an automatic rumor detecting algorithm. A CNN with a filter wrapper (information gain-ant colony) made up the model. At the model's output, rumor classification is performed using the naive Bayes classifier. In this experiment, two sets of features were extracted in this paper. First, textual features were extracted by CNN. The CNN used EL-Mo word vector model developed by [72] which generated word vectors from the context of the news by analyzing the correlation between the words, sentences, and documents. Next, optimal features were produced by the filter wrapper approach which used term frequency-inverse document frequency (tf-idf) which analyzed the importance of the keyword throughout the datasets. To train the classifier to determine if the news is true or false, both feature vectors are employed. To evaluate the performance of the model, the authors used PHEME datasets which contains events of five breaking news. The overall accuracy of this dataset was from 75% to 87%. By analyzing the context of both future and previous facts, the BiLSTM network was initially employed by Asghar et al. [6] to depict the long-term dependency in a news story. The CNN was then utilized to extract features and categorize whether or not the news is a rumor.

To identify rumors circulated during COVID-19 on social media platforms, Al-Sarem et al. [4] proposed a hybrid deep learning-based model that combines a long short-term

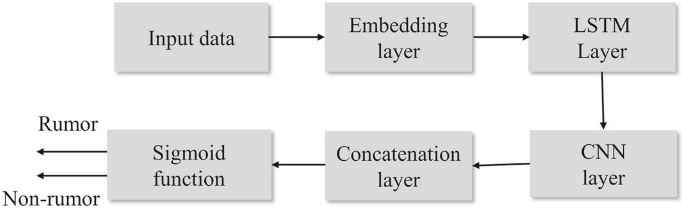


Fig. 9 LSTM-PCNN model

memory (LSTM) and concatenated parallel convolutional neural networks (PCNN). The longest tweets in the datasets are contained in the input layer. In embedding layers, the GloVe [71], Fast Text,[7] and word2vec [66] models are employed to develop word vectors. Word2vec use skip-gram and common bag of words where, as FastText uses n-grams to generate word embedding. The Glove is a model which build word vector by using word occurrences. The LSTM layer receives separate feeds from word embedding layers. To improve CNN performance and prevent the LSTM layer from overfitting, a dropout layer was added prior to the LSTM layer. Each CNN block produces a 150-dimensional vector in each of the three parallel CNN. After that, the output from each CNN was combined. In order to determine if a tweet contains rumors or not, a binary classification method using the sigmoid function was applied. The ArCOV-19 dataset, which includes Arabic tweets on the COVID-19 epidemic, was used by the authors. According to the experiment's findings, the LSTM-PCNN model beats the other baseline models and had the best results when the word2vec skip-gram model was applied. With a detection accuracy of 86.37%, it outperformed the other state-of-the-art models as well. The altered version of the original image referenced in [4] is shown in Fig. 9.

Azri et al. [8] presented an emotive lexicon-based model to capture the semantic characteristics and sentiments contained in tweets. The model is a hybrid one that searches microblogs for rumors using both CNN and LSTM. While long-term dependent information was recorded using LSTM, local features were extracted using the CNN. The input message's content, sentiments, and visual data were all examined by this model. Prior to sending all of this data for word embedding, it underwent preprocessing to create word vector. The VGG-19 model proposed by Simonyan and Zisserman [83] was used to extract features from visual data. CNN Text was applied to text data. Together, the text and sentiment vectors were supplied to the LSTM network. The next step was to apply the output of the LSTM and visual data to a binary classifier in order to determine if the tweet is a rumor or not. The datasets used for this experiment are from FakeNewsNet [82]. The authors created a dataset called DAT@Z20 using tweets they had collected from Twitter. For the FakeNewsNet dataset and the DAT@Z20 dataset, the model's accuracy was 94.3% and 92.2%, respectively. The altered version of the original image referenced in Azri et al. [8] is shown in Fig. 10.

The majority of the researchers, according to Shelke and Attar [79], used the text and temporal features of an input rumor post. As a result, the authors combined lexical features with user- and content-based features to create a hybrid model that combines the BiLSTM and multilayer perceptron (MLP) approaches. Word embedding was employed to create a word vector after preprocessing, which was then provided as input to the BiLSTM model, while the MLP model learned the post-wise features using user, lexical, and content-based features. Then, a densely linked layer that created a hybrid model was given a feature vector from an ensemble of MLP and BiLSTM models. The researchers used two datasets taken

[7] https://github.com/facebookresearch/fastText.

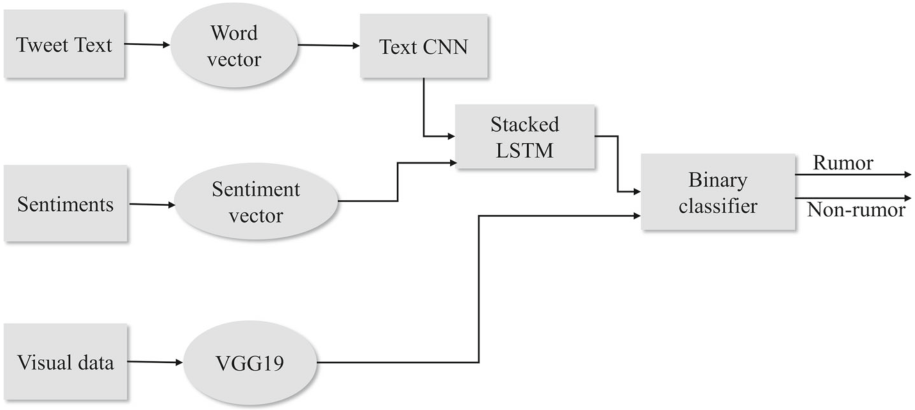


Fig. 10 deepMonitor model

from the real world and a benchmark dataset taken from Twitter to evaluate the performance of their approach. Two well-known websites, www.politifact.com and www.snopes.com, provided the real-world datasets, while Twitter provided the benchmark datasets. The author outperformed other models with great accuracy of 97%.

2.2.6 Other methods

Here, some attempts other than only DL-based techniques are discussed. In a short blog post, Ma et al. [62] investigated a framework for rumor detection that combines kernel learning with propagation trees to build higher-order sequences that distinguish between various rumor categories. In order to achieve more accuracy in rumor identification, Zhang et al. [110] developed a knowledge-aware network that makes use of numerous modal content and external knowledge-level relationships. The authors proposed a model that generates the post representation by treating the post's word, visual, and knowledge embeddings as multiple stacked channels similar to colored images while explicitly maintaining their alignment relationships to capture the post's full semantic meaning. In order to gather the event-independent latent topic information of events, extract the event-invariant features, and to enhance the capabilities of the rumor detection model, they also created the Event Memory Network (EMN) events' latent topic information. To train the model, the authors used both PHEME and Twitter datasets. The proposed model gave an accuracy of 81.6% for PHEME datasets and 86.6% for Twitter datasets. Ma and Gao [60] created a tree transformer strategy where user relationships such as tweets and retweets in terms of claims and the structure of the tree were employed to debunk rumors on social media. The contextual features from the input source were considered in this paper. Two publicly available benchmark datasets, Twitter and PHEME, were used in this work. Following the work of Ma et al. [63], the authors also used top-down and bottom-up approaches for this experiment. An attention layer was employed to obtain the correct data as well as the opinions prevalent across the entire tree. The authors also demonstrated that their suggested approach could identify rumors even while they were only starting to spread. They demonstrated that the proposed model had an F1 score of 65.9% on PHEME datasets while providing 75% accuracy on Twitter datasets.

According to [42], the proposed tree model of [51, 64] showed some limitations toward rumor detection. The authors mentioned that both the papers only focused on structural information present in the conversation thread. The information present in the conversation thread was from both the directions, i.e., from parent to child and vice versa. But the thread structure in social media conversation was not mentioned, i.e., their tree models did not mention any interaction between nodes from other branches. In order to overcome these limitations, Khoo et al. [42] et al. proposed a model where they focused on the community response within the social media conversation. They arranged all tweets in a chronological order. For this experiment, PHEME, Twitter 15 and Twitter 16 datasets were used. As the tweets were arranged linearly the author found that the model lost structural information. To overcome this problem they proposed few new variants of the model first they added self-attention mechanism. Further, for sentence representation token level self-attention was used before attention mechanism. And finally, they added time delay information for each tweet. Then, these new models were compared. The model with attention mechanism outperformed the other models and gave an accuracy of 85.2% for Twitter 15 and 87.4% for Twitter 16 datasets.

By examining content-based and context-based elements at three viewpoints, Jahanbakhsh-Nagadeh et al. [37] proposed a BERT-SAWS semi-supervised learning model for the early verification of Persian rumor. This model combined an unsupervised language representation with the text representation of the rumor's content with pre-trained bidirectional encoder representations from transformers—BERT. This text representation was employed for two things, including decreasing the issues with limited datasets in deep neural networks and early rumor verification. Lao et al. [52] created a double propagation model to identify rumor in the linear and nonlinear transmission fields. In this paper, the authors extracted some features like social context, content of the claim, and temporal information from the input data for detecting rumor. The claim content feature embedding and associated temporal state were completed initially. Then, utilizing various nonlinear rumor structures, the nonlinear diffusion properties were discovered. Then, sequential context data are merged to expand the representation of source nodes and generate distinctive vectors for sequential characteristics. Two datasets such as PHEME [114] and RumorEval [20] were used for this experiment. The proposed model gave an accuracy of 88.67% for PHEME datasets and 92.50% for RumorEval datasets. Dong et al. [23] proposed a two-step rumor detection methodology where different types of features were extracted and fed to different machine learning and DL models to classify rumor. Both user-based and non-user-based features fall under this category. The users or people who share the post are included in the user-based features. The substance of posts, diffusion-based features, and emotion-based features like joy, rage, and doubt are among the nonuser-based features. The fact that this work's features were based on super-network theory is crucial. The accuracy, precision, recall, and F1-score of the model were evaluated using four real-world themes, and the average results were 85.22%, 68.47%, 84.59%, and 75.14%, respectively.

Kotteti et al. [48] suggested an ensemble model (see Fig. 11) which consists of different deep neural network models. Preprocessing and ensemble modeling were the two stages used to explain the model. The input tweets were processed and converted into time series vectors in various time intervals during the preprocessing stage. Following data cleaning, data sparseness was reduced, data were normalized, and duplicate data were removed. After that, the model was fed the preprocessed data. The retweet counts were retrieved from the temporal features at each moment. Different basic learning neural networks, such as RNN, LSTM, and GRU, were used to build the ensemble model. PHEME datasets were utilized to assess the model's performance. The model's precision, recall, and F1 values of 64.3% were greater than those of the authors' earlier efforts [46, 47].

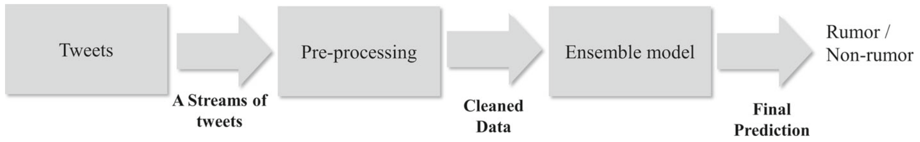


Fig. 11 Ensemble model

2.2.7 Discussion on the potential methods for rumor detection

We have covered a few models in this survey that have been identified as the best in terms of performance evaluation, technique, feature extraction, and classification results. We discovered that the majority of researchers created various DL-based models to identify rumors from microblogs [8]. A few of them also demonstrated how their algorithms were effective at spotting rumors before they spread [11, 64, 97, 100]. Some of them developed the models to extract features from both text and image data such as [8]. Again, researchers evaluated a variety of factors, including textual content, user profiles, and the transmission and dispersion structures of rumors, to determine if the communication is a rumor or not. A GCN-based tree-based recursive neural network model was proposed by Ma et al. [64], with an emphasis on the semantic sequence and propagation structure of the tweets in both directions. In contrast, Ma et al. [64] offered a GCN model that included post content and rumor dissemination structure to forecast the result. To find rumors in microblogs, Azri et al. [8] employed multimodal aspects of both textual and visual information. The majority of possible techniques demonstrated that their model predicts the output as rumor or non-rumor [4, 8, 14]. However, few researchers shown that their model predicts the output in four classifications, including non-rumor, false rumor, real rumor, and unconfirmed [106]. When addressing performance evaluation, Azri et al. [8] demonstrated that their model provided 94% accuracy, which was greater than other cutting-edge models, when they used a hybrid model of CNN and RNN for both text and image-based rumor data. In other paper, Shelke and Attar [79] showed that their model performed very well and gave an accuracy of 97% in two benchmark datasets collected from Twitter. They focused on lexical features and content-based features and developed a hybrid model based on BiLSTM and multilayer perceptron approaches.

2.3 Rumor tracking

Rumor tracking is a method for monitoring information and sifting through news that is rumor-related. Finding news stories linked to a certain incident is the main objective of rumor tracking [107]. When a rumor is discovered, the posts that are associated with it are gathered and filtered by tracking components [118].

In order to track rumors, Cheng et al. [17] created the VROC model, a multitask learning framework based on an RNN with BiLSTM and a variational autoencoder (VAE). A decoder and an encoder made up the VROC. To represent the information in the text in the compressed form, the encoder converted input tweets into tweet-level latent characteristics. To recreate the input text, the decoder decoded the features that the encoder had retrieved. The model independently completed four tasks such as rumor detection, rumor tracking, stance classification, and veracity classification producing unique loss functions for each one.

A deep reinforcement learning network-based ensembling model for rumor tracking was proposed by Li et al. [53]. The aim of this paper was to classify whether the collected tweets are related or not related to a given rumor. So, this paper used m-way classification for rumor tracking problem. The researchers used two datasets such as PHEME, RumorEval. The datasets contains a set of rumor events which contains a collection of tweet threads. They retrieved several content aspects and social information from the tweets, including “screen name,” “reply to screen name,” and “hashtag.” All the features were preprocessed to embed both the sets of tweets and rumor events into vector through several components. The naive Bayes, support vector machine, BiLSTM, FastText, and TextCNN models, among others, were first selected by the authors as components of the RLERT model for rumor tracking. Then all the components individually predicted the tracking output (related or not related). Finally, an ensembling technique was used by combining all the outputs of the components for a final prediction that if the tweets are related or not related to a rumor event.

In another work, Zeng and Cui [107] proposed rumor tracking classification whether the tweets are related or not to a particular event. The model was based on two presumptions, including the first that there may be connections between the tweets of the same events and the second that categorization was unable to discern between the tweets of various events because they were concealed from one another. The authors employed contrastive learning [109] along with these presumptions to cover the gaps in the tweets. Contrastive learning was primarily used to distinguish between two instances that were augmented from the same input source, allowing them to be closer in the representation space, and instances that were augmented from a separate source, allowing them to be farther away. Three steps were used to explain the model. Sentence-BERT, a pre-trained feature generator, was used to first map the tweets from text to feature space (SBERT). In order to produce positive and negative pairs based on the input source, contrastive learning was also applied. To learn the probability distribution of tweet information with respect to the event, three classification head variants—RNN-based, CNN-based, and fully connected neural network—were employed.

2.4 Stance classification

Rumor posture classification comes next after rumor detection and rumor monitoring. A lot of posts pertaining to the rumor are the result of the rumor tracking. According to [118], “stance classification” is the process of evaluating a tweet’s suitability for rumor veracity. As a result, stance classification is used to evaluate a news story’s sentiments or to measure a person’s attitude or emotion when they receive a piece of information, or even just their response to a tweet sentence [53, 102, 107]. There are basically four types of stances mentioned by the researchers according to different datasets. Xuan and Xia [102] explained four types of stance mentioned in RumorEval datasets such as support, deny, query, and comments.

- *Support* The respondent supports the authenticity of the rumor.
- *Deny* The respondent denies the authenticity of the rumor.
- *Query* Query refers to the time when the respondent will raise doubts about the truth and want additional evidence.
- *Comment* The respondents comment on the authenticity of the rumor.

The work of Lukasik et al. [58] had an influence on [74], who put forth a paradigm for stance classification. However, the model that the authors created was based on transfer learning and distinguished the unknown or concealed tweets from the known rumor tweets. By ignoring the rumor identities, the model by [74] concentrated on classifying tweets into rumor and non-rumor, and then tweets connected to rumor into support and not support. The authors represented the features using a bag of words (BOW).

By examining the four key categories of whether a tweet was supported or denied, or whether anyone queried or commented on its veracity, Zubiaga et al. [115] offered a methodology to address the stance classification of rumor. This paper analyzed tweets by considering the contextual features of the conversational threads. The conditional random field (CRF) classifier was utilized by the model to monitor the tweets and associated discussions. The conversions of tweets were used to create a graph using CRF, which was then used to create a series of stances. The authors used two different kinds of CRF, including linear CRF, which models each branch of the tree as a separate input to the classifier. Second, the Tree CRF has a collection of trees with complete tweet conversions as its input. The CRF classifiers demonstrated significant advancements for other classes, including supporting and querying tweets, where Tree CRF performs best.

The classification of the rumor's viewpoint is crucial for rumor verification. Therefore, Kochkina et al. [43] proposed a sequential network based on LSTM. In this experiment, the authors categorized tweets which they mainly supported, rejected, inquired about, and commented on rumor news. The authors described a unique strategy that makes use of the progression of changes seen in Twitter chat threads with a tree-like structure. They looked into the discussions that emerged from tweets that people were reacting to one another with. These responses lead to a tree-structured and frequently nested discussion, with each reply triggered by the originating tweet that started the conversation. Similar to this, Veyseh et al. [92] developed a CNN-LSTM model with deep attention that compiles a stream of input tweets. Additionally, the author included an attention mechanism that takes advantage of a variety of rumor stances. The experiment also revealed the tweets' feature relationships and tree structure.

Xuan and Xia [102] developed a framework for classifying rumor stances based on a number of features. The authors employed RumorEval datasets that focused on four stances: support, deny, query, and comments. They investigated 40 features offered by social media, and the three core features—text, user, and propagation features—were taken into consideration. They used the URL, topic, question, and exclamatory mark features in the text feature class to assess the type of stance. They have examined whether the user and source are verified, the number of followers, and other factors in the user feature class. And they included the number of likes and re-posts in the propagation feature class. The authors applied a variety of machine learning techniques based on these features, including logistic, naive Bayes, decision trees, and support vector machine algorithms. Then, classification algorithms were applied to those features. RumorEval was the dataset used for this research. The proposed model produced an accuracy of 80.5%. The proposed model showed good classification results for deny and query categories, and not satisfactory for other two categories.

2.5 Veracity classification

The purpose of veracity classification is to ensure that the user is receiving accurate information. When the phrase "classifying veracity of rumor" is used, it refers to determining the veracity of a rumor, according to which a rumor can either be true or untrue or not verified [17]. The rumor veracity classifier is used to identify whether a tweet is false, true, or unconfirmed after rumors have been detected and tracked. In order to classify the truth of rumors, Kumar and Carley [51] proposed a tree LSTM model that utilized convolution units. The original post and any associated comments were taken into consideration for this study. Each node in a tree created from the discussion represents a sentence. Three techniques, including branch LSTM, tree LSTM, and binary constituency tree LSTM, were used to learn the node representations. In branch LSTM, the branches of trees were utilized as input to the LSTM units, whereas in tree LSTM, the complete tree was used as input to the LSTM units. In order to reflect the inherent correlations in conversions in a certain task, the tree's structure was finally adjusted. The PHEME dataset with five events (see Sect. 3.1) served as the basis for this paper.

Rosenfeld et al. [76] looked into how information spreads and whether the legitimacy of unverified information spreading depends on how it spreads through social media. In order to extract topological information from the text data's structure, the authors used graph kernels. They developed a small number of models that are "sanitized" diffusion processes since they do not take into account the user, their language, or the time. The model correctly predicted whether the information would be true, false, or unverified. The model also performed admirably in the initial stages of information dissemination. In order to take advantage of the relationship between the tasks, Kochkina et al. [44] developed a multiple-task learning architecture that allows mixed training of both tasks in a verification pipeline. In order to improve the performance of veracity classification, the authors presented a rumor verification model that makes use of task relatedness with auxiliary tasks, such as rumor detection and stance classification. The proposed multitask learning methods combined the verification classifier with the stance and rumor detection classifiers independently, as well as with both the stance classifier and the rumor detection system, were contrasted with the single-task learning techniques. Similar to this, Roy et al. [77] proposed a DL model based on transformers that combined several encoders to display the co-relationship between the input. To capture various contextual interactions in latent feature space, the authors suggested a model named as globally discrete attention representation from transformers. The model used a variety of concurrent encoders in a way that allowed each encoder to discover a distinct association. The model also included a discrete attention component that helps to purge superfluous features from the feature space of different occurrences.

2.6 Rumor combating

Social interactions are mainly responsible for spreading rumors or other misinformation. So, it is important to develop some barriers to stop or slow down the process of propagation of rumor [29]. Rumor combating is the last phase of the rumor prevention process. Rumor combating is a method for preventing rumors from circulating on social media. Once the rumor was detected, tracked, and verified as rumor or non-rumor, the last step is to prevent the rumor from spreading further. Many researchers nowadays are trying to generate new models and methods to combat rumor in social media. Tripathy et al. [88] proposed two models for preventing rumors in social media. These were the delayed start model and the "beacon" model. In the first model, once a rumor is identified, a local authority begins to take steps to halt its spread. In contrast, in the second model only a few agents were assigned to the second method's mission to stop the spread of rumors. Once a rumor started to circulate on social media, they started disseminating counter-rumors to halt it. After identifying the rumor, any user can share the information with their neighbors in order to warn them and stop the rumor from spreading. This new model was proposed by Tripathy et al. [89] in order to improve the results of the previous work [88] on these two methods (delayed start model and beacon model). The fundamental distinction between these strategies is that any person can independently attempt to dispel rumors on social media. In this paper, the authors designed three models to spread some anti-rumors to prevent the spread of rumor in social media. Ji et al. [38] developed an anti-rumor dynamic model based on rumor dynamic theory to combat rumor in social media. The authors analyzed the model in a complex network by using mean field equations and some numerical simulations. The anti-rumor dynamics were characterized into three steps, i.e., the time at which anti-rumor should spread, the rumor propagation path, and finally, whether if the node accepts an anti-rumor as true or not. In this study, the authors employed two processes to examine how rumors and denials of rumors behave in complex networks. Anti-rumor priority over time (APOR) was the term used to describe the process where a node recognizes an anti-rumor as true. The process was known as prior hypothesis bias (PHB), and the condition was before accepting the anti-rumor, and it occurred when the node refused to accept the anti-rumor and gave priority to the rumor. The temporal threshold, a network-dependent variable that illustrates the dynamic nature of the anti-rumor spreading mechanism, was employed by the authors for additional investigation. The usefulness of the design concepts to counter health-related rumors in social media was tested by Öztürk et al. [70]. The experiment was carried out on a crowd-sourcing website[8] which is a business website where both the requester hire workers to do some tasks that computers could not do it. Zhao et al. [112] concentrated on a socio-psychology viewpoint to combat rumors in social media, in contrast to the work of [38, 88, 89], which only employed mathematical models. They claim that the current models only address people who actively sought to dispel rumors after becoming aware of them, and the amount of effort required is little. Zhao et al. [112] explained a few factors which particularly influenced the people's intention and their behavior to combat the rumors. The authors created a novel theory to describe how social media users make decisions to stop rumors by merging two theories, such as the norm activation model and theory of planned action. The idea of planned behavior described the incentive and informational influences on people's specific behaviors across a variety of disciplines. It was a rational choice theory. The norm activation model, on the other hand, was a pro-social behavior paradigm that clarified awareness of adverse effects, assigning blame, and personal standards.

3 Detailed analysis of rumor datasets, models' performance, and research gaps

This section focuses on the findings from the aforementioned survey on rumor detection, veracity and stance classification, tracking and combating which we analyze and discuss. Many researchers employed diverse datasets and various strategies to isolate rumors in social media, demonstrating the accuracy of their models by contrasting their results with other cutting-edge techniques. The biggest obstacle is gathering social media messages or postings and categorizing them as rumors or not rumors in order to create a usable dataset. While certain researchers went to great lengths to build annotated datasets for rumor detection, veracity and stance classification, and other purposes, others just utilized these datasets when they were made accessible to the public. To begin with, the several well-known datasets that have been utilized for DL-based rumor detection models by researchers in the past are summarized and explained in this section (see Table 1 and Sect. 3.1). We also made an effort to demonstrate many performance metrics used by researchers to evaluate the effectiveness of their methodologies (see Sect. 3.2). We have more thoroughly contrasted their research and models' performance on diverse datasets with other reference models (see Tables 2 and 3). Last but not least, We assessed, compared, and contrasted the potential research works critically, emphasizing the challenges they encountered in identifying potential research gaps (see Sect. 3.3).

[8] https://www.mturk.com/mturk/.

3.1 Discussions on rumor datasets

Collecting usable datasets from social media platforms that can be used for rumor detection remains a significant difficulty [69]. When collecting data (messages or posts) about various events, researchers frequently use fact-checking websites or social media platforms like Twitter. They cleaned and filtered the data using NLP-based preprocessing techniques. They then annotate the data and occasionally extract features before outputting a label or class for rumor, non-rumor, etc. When these datasets are made public, other researchers directly use them. In this section, we go over the datasets created and applied by different researchers for rumor detection, veracity classification, etc. Up till now, PHEME, Weibo, and Twitter datasets have been the most frequently used datasets [87]. Other datasets, like FakeNewsNet, DAT@Z20 [8], and ArCOV-19 [4], are also used to detect rumors. Twitter 15 datasets were developed by [62] using [61] and [56]'s as a guide. Twitter and Weibo datasets were created by [61] and [62]. They enlisted the aid of www.snope.com and the Sina community management center[9] to verify rumors and non-rumors. They also included some events from publicly accessible databases in order to balance the datasets. They ultimately received 992 total events, of which 498 were rumor datasets and 494 were non-rumor datasets.

The PHEME dataset, which [114] prepared, contains a collection of Twitter news posts on rumors and non-rumors made during breaking news. The author has gathered tweets from the Twitter API account while examining two distinct situations: first, rumors that are presented as breaking news, and second, specific rumors that have been pre-identified, then gathered tweets on nine different events, such as Sydney siege, Charlie Hebdo, Ottawa shooting, Ferguson, Prince to play in Toronto, Gurlitt, Germanwings Crash, Prince to play in Toronto, and Gurlitt, five of which were breaking news and four of which were particular news, and they marked each tweet with its veracity value, which might be true, false, or unverified. Azri et al. [8] built DAT@Z20 dataset by gathering data from Twitter and compiling it into a single file. The dataset contains 8,999 news articles of only data and metadata. The authors used Twitter API to retrieve the surrounding social context of the particular tweets, such as replies, and retweets. They have collected a total of 2,496,982 tweets, from which only kept the tweets containing both text and visual data and thus removed only text and duplicate image data. So, the total tweets of 249,076 had taken into consideration, which they further split into train and test data for the experiment. The authors label the datasets as fake or true.

Another dataset—FakeNewsNet, was initially developed by [82], which contains fake and legitimate news stories collected from the fact-checking websites like PolitiFact[10] and Gossip Cop. Azri et al. [8] used this dataset for their experiment to classify both fake news with images. Therefore, the authors removed all the irrelevant tweets, such as tweets with images which are duplicates and of low quality. Thus, from a total of 1,607,760 tweets, only 207,768 tweets were used for the experiment. With all required features and descriptions, Table 1 highlights some notable datasets used for rumor detection or classification.

3.2 Performance evaluation criteria

To assess the performance of the models, researchers contrasted them with other state-of-the-art models. They typically provided a confusion matrix, accuracy, precision, recall, etc., to demonstrate how well their proposed method was explained. The confusion matrix is a table that evaluates the classification model's performance in correctly categorizing instances into

[9] http://service.account.weibo.com.

[10] https://www.politifact.com/.

Table 1 Significant datasets on rumor detection

Datasets	Total samples	Output label	Distribution of output labels	Description
Dat@z20 [8]	8999 (news) 1313 (news with image data)	Fake/true	6496/2503 (news data) 858/455 (news with image data)	The datasets contain both news articles as well as image data
Weibo [61]	4664	Rumor/non-rumor	2313/2351	The datasets are collected from the Sina community management center
PHEME (version 1) [114]	7,507	Rumor/non-rumor	2695/4812	This dataset contains 9 events which includes (rumor/non-rumor) Sydney siege -535/786, Charlie Hebdo 474/1695, Ottawa shooting 474/426, Ferguson 291/892, Germanwings Crash 331/690,Prince to play in Toronto 237/4, Gurlitt 190/196,Putin missing 143/123, Essien has Ebola 18/0
PHEME (version 2) [117]	5,802	Rumor/non-rumor	1972/3830	This dataset contains 5 events which includes(rumor/non-rumor) Sydney siege -522/699, Charlie Hebdo 458/1621, Ottawa shooting 470/420, Ferguson 284/859, Germanwings Crash 238/231
Twitter 15 [62]	1490	False/True/ Unverified/ non-rumor	370/374/374/372	These datasets have 276,663 users, 1490 source tweets, 331,612 threads
Twitter 16 [62]	818	Rumor/non-rumor	205/205/203/205	This dataset contains 173,486 users, 818 source tweets and 204820 threads
ArCov19 [4]	157	Rumor/non-rumor	1480/1677	The ArCOV-19 dataset contains Arabic tweets described the COVID-19 pandemic
RumorEval [20]	4618	Rumor/non-rumor	325	This dataset contains 4017 branches from Twitter
Gossip Cop[82]	1,043	Rumor/non-rumor	619/424	The dataset was obtained from the FakeNewsNet
FAKENEWSNET [82]	23196 (news articles) 19200 (news with image)	Fake/true	5755/17,441 (news articles)1986/17214 (news with image)	True and false news articles are collected from the fact-checking websites

various groups [30]. Accuracy is defined as the ratio of the number of messages successfully classified as rumors to all messages. And it is denoted as Eq. 1.

$$\mathrm{Accuracy(A)} = \mathrm{(TP + TN)/(TP + FP + TN + FN)} \tag{1}$$

where TP= True positive, TN= True negative, FP=False positive, and FN=False negative. Precision is calculated (see Eq. 2) as the ratio of all messages correctly classified as rumors (TP) to all messages classified as rumors (TP + FP).

$$\mathrm{Precision(P)} = \mathrm{TP/(TP + FP)} \tag{2}$$

Recall is the ratio (see Eq. 3) of all messages correctly classified as rumors (TP) to all messages that should be classified as rumors (TP + FN).

$$\mathrm{Recall(R)} = \mathrm{TP/(TP + FN)} \tag{3}$$

F1-measure is the harmonic mean of precision and recall (see Eq. 4).

$$F1 = \mathrm{2PR/(P + R)} \tag{4}$$

3.3 Critical analysis with an overview of the performances of DL-based models on standard datasets

We conducted the detail review and found that researchers created numerous DL-based rumor detection models and assessed their performances on various datasets. In order to assess some of the remarkable systems' accuracy, precision, recall, overall system performance, and uniqueness, we critically analyzed them. Along with this, we have given a comparison between DL-based methods and the performance of the models on standard datasets. Table 2 shows an overview of the various models we surveyed and described in Sect. 2 and their performances. Table 3 compares the accuracy of all DL-based models grouped by the most often used datasets. We discovered that the majority of researchers trained their models using the Twitter 15 and Twitter 16 datasets.

In this survey, as we said before, we analyzed the DL-based system models for rumor detection. The researchers extracted various features to improve the performance of their model of detecting rumors in social media. Among them, few researchers showed that their model were best in terms of accuracy, precision, recall, and F1 value. For example, Ma et al. [61] developed a model based on RNN to capture the temporal features present in the tweets and got an accuracy of 91% on Weibo datasets. However, using same Weibo dataset Liu et al. [57] compared their work to [61] and mentioned that Ma et al. [61] only used forwarding comments and neglected the dynamic behavior of the spreaders and diffusion structures of rumors. Therefore, Liu et al. [57] not only focused on the forwarding contents but also spreaders and diffusion structure of messages to detect rumor. They proved that their model gave better results in terms of accuracy (94.4%) using LSTM along with a pooling layer of CNN as compared to [61] who used only RNN. On the contrary, with the same dataset, Yang et al. [105] used a GCN to study the relationship between post and comments and compared their work with the model by [11]. According to the researchers, Bian et al. [11] only focused on propagation structure of rumor by analyzing only structure of repost content and its structure. Bian et al. [11] also mentioned that collecting huge repost structures from real-world social networking sites is challenging and expensive due to which sufficient information may not be available. Yang et al. [105] have proved that their model gave good results as compared to the model developed by [11]. To verify this, they further experimented with the model

Table 2 significant DL models comparison on standard dataset

Year	DL models applied	Datasets	Performance
2016	RNN [61]	Twitter, Sina Weibo	Accuracy−91.0% (Twitter) Accuracy−88.1% (Weibo)
2018	RNN [15]	Twitter, Weibo	Precision−74.02% (Twitter), 71.7% (Weibo), Recall−68.75% (Twitter), 70.34% (Weibo)
2019	BiLSTM, CNN[6]	PHEME	Accuracy−86.1%
2019	LSTM, CNN[57]	Sina Weibo	Accuracy–94% precision-93% recall-95% F1-94%
2019	GCN [33]	Twitter 15, Twitter 16	Accuracy77.3% (Twitter 15) Accuracy−75.2% (Twitter 16)
2019	RNN [54]	RumorEval, PHEME	Accuracy−63.8% (RumorEval) Accuracy−48.3% (PHEME)
2020	BiLSTM [86]	PHEME	Accuracy−92.6%(PHEME 2017) Accuracy−91.9% (PHEME2018)
2020	GCN [11]	Sina Weibo, Twitter 15, Twitter 16	Accuracy−96.1% (Sina Weibo) Accuracy−88.6% (Twitter 15) Accuracy−88.0% (Twitter 16)
2020	CNN [27]	YELP-2, FBN	Accuracy−87.2% precision−79.1% recall−84.7% F1-82%
2020	Deep learning [48]	PHEME	Accuracy−94.9% precision−37.4% recall−51.8% F1-79%
2020	Deep learning, CNN [5]	PHEME	Accuracy–94%
2021	Deep learning, LSTM [94]	PHEME	Accuracy–81% precision-79% recall-79% F1-81%
2021	GCN [93]	Twitter 15, Twitter 16	Accuracy−87.8% (Twitter 16) Accuracy−85.6% (Twitter 15)
2021	Propagation structure, CNN [90]	Twitter 15, Twitter 16, Weibo	Accuracy−95.1% precision−94.5% recall−95.6% F1−95.0%
2021	GCN, [97]	PHEME, Twitter 15, Twitter 16	Accuracy–89% (Twitter 15) Accuracy−91.5% (Twitter 16) Accuracy-69% (PHEME)
2021	GCN [9]	PHEME	Accuracy−84.1% Precision−88.2% Recall−95.6% F1- 89.6%
2021	Naive Bayes classifier [50]	PHEME	Accuracy−76.7%, precision−76.1% recall−76.3%
2021	Deep neural networks [8]	DAT@Z20, Fake News Net	Accuracy-94% precision-93% recall-95%
2021	CNN, LSTM [4]	ArCOV-19	Accuracy–85% precision-85% recall-85% F1-85%
2022	GCN [106]	Twitter 15, Twitter 16	Accuracy−86.5% (Twitter 16) Accuracy−83.6% (Twitter 15)
2022	BiLSTM [14]	Weibo	Accuracy-95% precision−94.3% recall−94.1%

Table 3 Performance of DL-based models on standard datasets

Datasets	DL-based models	Accuracy	Year of publication
Twitter 15	RvNN [64]	72.3%	2018
	Hybrid GCN [33]	75.2%	2019
	Bi-GCN [11]	88.6%	2020
	Rumor2vec [90]	79.6%	2021
	EB-GCN [97]	89.2%	2021
	RDGCN[93]	85.6%	2021
	HDGCN [106]	83.4%	2022
	GAN [100]	85.6%	2022
Twitter 16	RvNN[64]	73.7%	2018
	Hybrid GCN[33]	77.3%	2019
	Bi-GCN [11]	88.0%	2020
	Rumor2vec [90]	85.2%	2021
	EB-GCN [97]	91.5%	2021
	RDGCN [93]	87.8%	2021
	HDGCN [106]	86.5%	2022
	GAN [100]	85.4%	2022
PHEME	RNN [54]	63.8%	2019
	ERD model [94]	84%	2021
	EB-GCN [97]	69.0%	2021
	EGCN [9]	84.1%	2021
Weibo	RNN [61]	91.0%	2016
	RNN [57]	94.8%	2018
	Bi-GCN [11]	96%	2020
	Postcom2vec [105]	95.41%	2021
	Rumor2vec [90]	95.1%	2021
	GAN [100]	94.1%	2022

proposed by [11] to check the effectiveness and found that their model outperformed [11] by 2.99%. Few other researchers used CNN and GCN models to use propagation structures of rumors tweets [11, 64, 90, 97]. However, these researchers worked on same datasets like Twitter 15, Twitter 16, PHEME, and Weibo datasets. Tu et al. [90] used a CNN-based model to capture textual contents of source tweets and propagation structures of different tweets in a single graph which helped them to achieve an accuracy of 79.6% for Twitter 15 and 85.2% for Twitter 16 datasets.

In other hand, researchers like [11, 64, 97] used GCN-based models to study propagation structures of rumor in social media. Ma et al. [64] used the hidden representations of propagation structures of tweets and focused on recursive neural network and showed that their model was able to produce 72.3% accuracy on Twitter 15 and 73.3% on Twitter 16 datasets. Looking at this work, Bian et al. [11] focused on both propagation and dispersion structure of rumors and used a GCN model and showed that their model outperformed the model developed by [64] and generated 88.6% accuracy on datasets. In order to move forward this work once again [97] argued with the work done by [64] and [11] and showed that these two researchers only concentrated on text and propagation structure of rumor and, however, missed the uncertainty occurred in propagation structures. Therefore, in order to analyze this property the authors developed an edge enhanced Bayesian graph on GCN model and, finally, demonstrated that their model was able to produce an accuracy of 89.2% for Twitter 15 and 91.5% on Twitter 16 datasets (the best results so far as given in Table 3). After reviewing the above papers [11, 64, 97], Wang et al. [93] explained that those papers only focused on the propagation structure of rumor and, however, neglected the transmission pattern. So, the authors used a GCN model to analyze regionalized rumor transmission pattern. To address the over-smoothing issue which was ignored by the above papers and enhance the model's learning capacity, the author introduced a source text-enhanced residual GCN layer to the mode. They showed that their model gave best results as compared to the above papers in terms of accuracy. To verify this they considered the above works as baseline and further experimented on that. The author mentioned that the above models produce accuracy of 72.3% [64], 81.9% [11] and 82.5% [97] on Twitter 15 datasets where as their model gave 85.6% of accuracy on same dataset. Further, [106] noticed that no one from the above work focused on the dynamic behavior of the rumor. In order to explore the dynamic propagation of rumor, the authors proposed a GCN-based model and compared their work to other researchers like [64] and,[90]. They achieved an accuracy of 83.4% and 86% on same datasets like Twitter 15 and Twitter 16. However, this work can be further analyzed by exploring both dynamic propagation and dispersion of rumor.

Some researchers used both RNN and CNN to develop a model for rumor detection. According to them, to improve the performance, an ensemble of RNN and CNN should be used. For example, Azri et al. [8] used a hybrid model using LSTM and CNN to explore text and image data along with sentiments. According to them, the amount of work in these areas is quite low and hence their model also outperformed the other models and produced an accuracy of 94.3% and 92.2% on Fake News net and DAT@20 datasets. In another work, Kumar et al. [50] used a hybrid model to explore the textual features and to train their model took PHEME datasets. Hence, the researcher produced 73.2% F1 value as compared to other models. In another work, [79] used BiLSTM network with a multilayer perceptron and extracts user, text, and content-based features along with lexical features and produced 97% of accuracy on real-world dataset collected from Twitter. To conclude with the critical analysis, we have seen that GCN-based models proposed by different researchers performed best on the bench-marked rumor detection datasets with respect to other DL-based models. Hybrid ensemble models are also performing well on the rumor datasets for rumor detection.

To conclude the critical analysis, we can say that GCN-based models, as proposed by several researchers, outperformed other DL-based models on benchmark rumor detection datasets. On the rumor datasets, hybrid ensemble models are also effective in detecting rumors. Since there are not many notable publications on rumor tracking, veracity, and stance classification in the literature, we chose not to analyze these system models. The challenges and potential gaps in this field of research are listed in the next subsection.

3.4 Research gaps and challenges

Researchers from all over the world have worked very hard to improve rumor detection and prevention systems in recent years. However, there are still several areas that require investigation. The tasks of rumor detection and veracity classification received the majority of the researchers' attention. This subsection summarizes in-depth research carried out by knowledgeable researchers and projects the challenges encountered by them in rumor detection, tracking, and combating, allowing for the investigation of potential research gaps.

We came to the conclusion that DL-based models have contributed to more effective rumor detection after conducting a thorough survey on the process of rumor detection techniques in social media. To summarize, the list below shows the major challenges in the field of rumor detection and combating.

- *Lack of good annotated rumor datasets with features* There were less publicly accessible benchmark datasets. There are other social networking sites like Facebook, WhatsApp, and Instagram in addition to Twitter. Therefore, more fresh datasets from these websites should be investigated. Images and videos can propagate rumors as well, and there were less number of image and video datasets available.
- *Challenges in rumor detection* According to Yang et al. [105], their model failed to select the interpretable comments for the rumor detection output. Similar to this, Asghar et al. [6] noted two significant weaknesses in their model. First, their model only considers text-based features; it excludes all other features, and second, their model only applies to English text. The most challenging thing during rumor propagation is to detect the rumor at the time of dissemination which is known as early rumor detection, because at the beginning stage it is quite difficult to identify a news to be rumor or not as the amount of information which is propagated will be less. Many researchers worked on early rumor detection by developing different DL-based models [11, 57, 59, 64, 93, 97, 105]. These authors mainly focused on different types of GCN based on top-down and bottom-up approaches to detect rumor at its early stage. All these authors have basically used the same Twitter 15 and Twitter 16 datasets. But, the amount of work on detecting the rumor at its early stage is quiet less.
- *Challenges in rumor tracking* Amount of work on rumor tracking is quite less due to the lack of datasets. Further, the datasets used by researchers for rumor tracking were either small in size or mostly imbalanced.
- *Challenges in rumor combating* Rumor combating in social media is a biggest challenge nowadays. After detecting and tracking the rumor, it is very much important to prevent those rumors from further spread. As per our survey, very few papers focused on developing some mathematical models to combat rumor in social media [88, 89]. The amount of work in developing a model to combat rumor by using DL technique is very less. Choi et al. [19] developed a DL-based model to combat rumor in social media by taking a claim sentence of a rumor. However, their model only focused on English language and failed to work on non-English speaking countries. To combat rumor developing a strategy to locate source of the rumor, understand it, and then act quickly to address, it is the biggest challenge nowadays in social media.

Next, we have highlighted some of the most significant research gaps that have not yet been adequately addressed by researchers. Some of them are described in the list below.

- *Gaps in rumor detection* We found that the existing work can be enhanced by adding more features and lexicon-based sentiment analysis for autonomous rumor detection. Moreover, some researchers produced models that are only available in a certain language like English. Advanced NLP with the most recent deep neural network has a lot of potential for improving the accuracy of existing rumor detection models, incorporating explanatory remarks for rumor detection findings, and expanding rumor detection models to new languages. In case of rumor detection, the most important things are the features that can help to decide whether the tweet is a rumor or not and sometimes DL classifiers failed to extract those features. Therefore, in order to deal with the context of the tweets, unique features should be extracted and applied. Additionally, there is relatively little work involved in detecting rumors using unsupervised and semi-supervised learning.

- *Gaps in rumor dataset preparation* The amount of datasets that can be used for rumor detection or any other activity is extremely little. The majority of the datasets that the researchers used in their model were obtained from Twitter. However, new datasets should be created using data from Facebook, Instagram, and other social media sites. Again, based on our survey, we discovered that there are very little image and video datasets available. However, as fraudulent images and videos proliferate in the digital age, it is crucial to create image and video datasets or images with text data. Additionally, rumors are shared in a variety of languages, and therefore, creating datasets in additional languages should be the focus of future research.
- *Gaps in rumor tracking* Larger datasets should be created in order to train the model because there are not enough datasets for rumor tracking. It is also important to extract some distinguishing features, including text similarity, to determine how closely the posts related to an event, in order to track rumors on social media. Additional applications of linked-graph, semantics-based rumor networks related to particular types of events, etc., are possible.
- *Gaps in rumor combating* Most researchers who have contributed to the rumor-combating process, focused on constructing mathematical models to study the mechanisms of rumor suppression. In order to regulate rumors, these models largely focus on the network's structural characteristics. More potent factors that can influence rumors include "trust" between the people involved and the rumor's context. By taking these elements into account, it will be interesting to analyze how rumors are combated in real-world data. And also, implementing innovative rumor-busting strategies, such as rumor bots, server applications to block rumors, and soft strategies like early rumor marking, applying cybersecurity strategies will stifle rumors at or near their source.

4 Conclusion

With the advent of social media, the pace of information dissemination has increased exponentially. There are no geographical boundaries, ensuring availability of information across the globe with lightning speed. This has enabled us to access plethora of information spanning around multiple fields. It helps us immensely in judicious decision making. This also opens the gate for dissemination of misinformation or rumor masqueraded as veracious information. The objective of this survey paper is to give a brief overview of modern advancements in rumor detection and prevention methods. In this survey, we reviewed many recent research works on rumor detection, rumor veracity and stance classification, rumor tracking and combating on social networking sites. We have discussed briefly on misinformation and its related terms such as fake news and rumor. We have examined the most recent DL-based methods for rumor detection, contrasted them, and compared how well they performed. We have also analyzed different methods related to rumor stance and veracity classification, rumor tracking, and rumor combating. Finally, we have done a critical analysis of all remarkable works highlighting research gaps and challenges that might be possible subjects for further study. In this survey, a detailed summary of all the standard rumor datasets and their sources are provided. On each of the datasets, we additionally examined the accuracy of the state-of-the-art DL-based methods in order to compare them. Further, while doing this survey we missed few other kinds of approaches used for rumor detection such as unsupervised and semi-supervised learning methods. We should include additional research on rumor tracking and combating in social media, which are not as widely available, in order to improve this survey. There is scope of improvement of this survey by exploring other type of misinformation and comparing them to rumors in terms of features, detection methods, and datasets used.

Author Contributions Barsha Pattanaik surveyed related papers thoroughly in this domain. Barsha Pattanaik and Sourav Mandal wrote the main manuscript text. Rudra M. Tripathy reviewed the manuscript.

Data availability Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study. The datasets we mentioned in this article are properly cited.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this research paper.

References

1. Ahmad T, Faisal MS, Rizwan A, et al (2022) Efficient fake news detection mechanism using enhanced deep learning model. Appl Sci 12(3). https://www.mdpi.com/2076-3417/12/3/1743
2. Ahmed H, Traoré I, Saad S (2017) Detection of online fake news using n-gram analysis and machine learning techniques. In: Intelligent, secure, and dependable systems in distributed and cloud environments - first international conference, ISDDC 2017, Vancouver, BC, Canada, October 26–28, 2017, Proceedings, Lecture Notes in Computer Science, vol 10618. Springer, pp 127–138, https://doi.org/10.1007/978-3-319-69155-8_9
3. Al-Sarem M, Boulila W, Al-Harby M et al (2019) Deep learning-based rumor detection on microblogging platforms: a systematic review. IEEE Access 7:152,788-152,812. https://doi.org/10.1109/ACCESS.2019.2947855
4. Al-Sarem M, Alsaeedi A, Saeed F, et al (2021) A novel hybrid deep learning model for detecting covid-19-related rumors on social media based on lstm and concatenated parallel cnns. Appl Sci 11(17). https://doi.org/10.3390/app11177940, https://www.mdpi.com/2076-3417/11/17/7940
5. Alsaeedi A, Al-Sarem M (2020) Detecting rumors on social media based on a cnn deep learning technique. Arab J Sci Eng 45(12):10,813-10,844
6. Asghar MZ, Habib A, Habib A et al (2021) Exploring deep neural networks for rumor detection. J Ambient Intell Humaniz Comput 12(4):4315–4333. https://doi.org/10.1007/s12652-019-01527-4
7. Aslam N, Khan IU, Alotaibi FS et al (2021) Fake detect: a deep learning ensemble model for fake news detection. CompLex 2021:5557,784:1-5557,784:8. https://doi.org/10.1155/2021/5557784
8. Azri A, Favre C, Harbi N, et al (2021) Calling to CNN-LSTM for rumor detection: A deep multi-channel model for message veracity classification in microblogs. In: Machine learning and knowledge discovery in databases. Applied data science track - European Conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, Proceedings, Part V, Lecture Notes in Computer Science, vol 12979. Springer, pp 497–513, https://doi.org/10.1007/978-3-030-86517-7_31
9. Bai N, Meng F, Rui X et al (2021) Rumour detection based on graph convolutional neural net. IEEE Access 9:21,686-21,693. https://doi.org/10.1109/ACCESS.2021.3050563
10. de Beer D, Matthee MC (2020) Approaches to identify fake news: a systematic literature review. Integr Sci Digital Age 136:13–22
11. Bian T, Xiao X, Xu T, et al (2020) Rumor detection on social media with bi-directional graph convolutional networks. In: The thirty-fourth AAAI conference on artificial intelligence, AAAI 2020, the thirty-second innovative applications of artificial intelligence conference, IAAI 2020, the tenth AAAI symposium on educational advances in artificial intelligence, EAAI 2020, New York, NY, USA, February 7–12, 2020. AAAI Press, pp 549–556, https://ojs.aaai.org/index.php/AAAI/article/view/5393
12. Bondielli A, Marcelloni F (2019) A survey on fake news and rumour detection techniques. Inf Sci 497:38–55. https://www.sciencedirect.com/science/article/pii/S0020025519304372
13. Celliers M, Hattingh M (2020) A systematic review on fake news themes reported in literature. In: Responsible design, implementation and use of information and communication technology - 19th IFIP WG 6.11 conference on e-Business, e-Services, and e-Society, I3E 2020, Skukuza, South Africa, April 6–8, 2020, Proceedings, Part II, Lecture Notes in Computer Science, vol 12067. Springer, pp 223–234, https://doi.org/10.1007/978-3-030-45002-1_19

14. Cen J, Li Y (2022) A rumor detection method from social network based on deep learning in big data environment. Comput Intell Neurosci 2022
15. Chen T, Li X, Yin H, et al (2018) Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection. In: Trends and applications in knowledge discovery and data mining - PAKDD 2018 workshops, BDASC, BDM, ML4Cyber, PAISI, DaMEMO, Melbourne, VIC, Australia, June 3, 2018, Revised Selected Papers, Lecture Notes in Computer Science, vol 11154. Springer, pp 40–52, https://doi.org/10.1007/978-3-030-04503-6_4
16. Chen X, Wang C, Li D, et al (2021) A new early rumor detection model based on bigru neural network. Discrete Dyn Nat Soc
17. Cheng M, Nazarian S, Bogdan P (2020) Vroc: Variational autoencoder-aided multi-task rumor classifier based on text. In: WWW '20: the web conference 2020, Taipei, Taiwan, April 20–24, 2020. ACM/IW3C2, pp 2892–2898, https://doi.org/10.1145/3366423.3380054
18. Cho K, van Merrienboer B, Bahdanau D, et al (2014) On the properties of neural machine translation: Encoder-decoder approaches. In: Proceedings of SSST@EMNLP 2014, eighth workshop on syntax, semantics and structure in statistical translation, Doha, Qatar, 25 October 2014. Association for computational linguistics, pp 103–111, https://aclanthology.org/W14-4012/
19. Choi D, Oh H, Chun S et al (2022) Preventing rumor spread with deep learning. Expert Syst Appl 197(116):688. https://doi.org/10.1016/j.eswa.2022.116688
20. Derczynski L, Bontcheva K, Liakata M, et al (2017) Semeval-2017 task 8: Rumoureval: determining rumour veracity and support for rumours. In: Proceedings of the 11th international workshop on semantic evaluation, SemEval@ACL 2017, Vancouver, Canada, August 3–4, 2017. Association for Computational Linguistics, pp 69–76, https://doi.org/10.18653/v1/S17-2006
21. Boididou C, Papadopoulos S, Zampoglou M, Apostolidis L, Papadopoulou O, Kompatsiaris Y (2018) Detection and visualization of misleading content on Twitter. Int J Multimedia Inf Retrieval 7(1):71–86. https://doi.org/10.1007/s13735-017-0143-x
22. Devlin J, Chang M, Lee K, et al (2019) BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 conference of the north american chapter of the association for computational linguistics: human language technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2–7, 2019, Volume 1 (Long and Short Papers). Association for Computational Linguistics, pp 4171–4186, https://doi.org/10.18653/v1/n19-1423
23. Dong X, Lian Y, Chi Y et al (2021) A two-step rumor detection model based on the supernetwork theory about weibo. J Supercomput 77(10):12,050-12,074. https://doi.org/10.1007/s11227-021-03748-x
24. Fernández M, Alani H (2018) Online misinformation: challenges and future directions. In: Companion of the the web conference 2018 on the web conference 2018, WWW 2018, Lyon , France, April 23–27, 2018. ACM, pp 595–602, https://doi.org/10.1145/3184558.3188730
25. Gaur L, Arora GK, Jhanjhi NZ (2022a) Deep learning techniques for creation of deepfakes. In: DeepFakes. CRC Press, pp 23–34
26. Gaur L, Mallik S, Jhanjhi NZ (2022b) Introduction to deepfake technologies. In: DeepFakes. CRC Press, Boca Raton, pp 1–8
27. Guo M, Xu Z, Liu L, et al (2020) An adaptive deep transfer learning model for rumor detection without sufficient identified rumors. Math Probl Eng 2020
28. Gupta A, Lamba H, Kumaraguru P, et al (2013) Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In: 22nd international world wide web conference, WWW '13, Rio de Janeiro, Brazil, May 13–17, 2013, companion volume. international world wide web conferences steering committee/ACM, pp 729–736, https://doi.org/10.1145/2487788.2488033
29. Habiba, Yu Y, Berger-Wolf TY, et al (2008) Finding spread blockers in dynamic networks. In: Advances in social network mining and analysis, second international workshop, SNAKDD 2008, Las Vegas, NV, USA, August 24–27, 2008, revised selected papers, lecture notes in computer science, vol 5498. Springer, pp 55–76, https://doi.org/10.1007/978-3-642-14929-0_4
30. Han J, Kamber M (2006) Data mining: concepts and techniques. Second Edition. The Morgan Kaufmann series in data management systems, Elsevier
31. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
32. Huang MZ, Yin RW (2022) Application research of fake news and rumors detection in complex network environment. Math Probl Eng
33. Huang Q, Zhou C, Wu J, et al (2019) Deep structure learning for rumor detection on twitter. In: International joint conference on neural networks, IJCNN 2019 Budapest, Hungary, July 14–19, 2019. IEEE, pp 1–8, https://doi.org/10.1109/IJCNN.2019.8852468

34. Huang Y, Chen P (2020) Fake news detection using an ensemble learning model based on self-adaptive harmony search algorithms. Expert Syst Appl 159(113):584. https://doi.org/10.1016/j.eswa.2020.113584
35. Islam MR, Liu S, Wang X et al (2020) Deep learning for misinformation detection on online social networks: a survey and new perspectives. Soc Netw Anal Min 10(1):82. https://doi.org/10.1007/s13278-020-00696-x
36. Islam N, Shaikh A, Qaiser A et al (2021) Ternion: an autonomous model for fake news detection. Appl Sci 11:9292. https://doi.org/10.3390/app11199292
37. Jahanbakhsh-Nagadeh Z, Feizi-Derakhshi M, Sharifi A (2021) A semi-supervised model for persian rumor verification based on content information. Multim Tools Appl 80(28–29):35,267-35,295. https://doi.org/10.1007/s11042-020-10077-3
38. Ji K, Liu J, Xiang G (2014) Anti-rumor dynamics and emergence of the timing threshold on complex network. Phys A 411:87–94
39. Jin Z, Cao J, Guo H, et al (2017) Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In: Proceedings of the 2017 ACM on multimedia conference, MM 2017, Mountain View, CA, USA, October 23–27, 2017. ACM, pp 795–816, https://doi.org/10.1145/3123266.3123454
40. Jwa H, Oh D, Park K, et al (2019) exbake: Automatic fake news detection model based on bidirectional encoder representations from transformers (bert). Appl Sci 9(19). https://www.mdpi.com/2076-3417/9/19/4062
41. Kaliyar RK, Goswami A, Narang P (2021) Fakebert: Fake news detection in social media with a bert-based deep learning approach. Multim Tools Appl 80(8):11,765-11,788. https://doi.org/10.1007/s11042-020-10183-2
42. Khoo LMS, Chieu HL, Qian Z, et al (2020) Interpretable rumor detection in microblogs by attending to user interactions. In: The thirty-fourth AAAI conference on artificial intelligence, AAAI 2020, the thirty-second innovative applications of artificial intelligence conference, IAAI 2020, the tenth AAAI Symposium on educational advances in artificial intelligence, EAAI 2020, New York, NY, USA, February 7–12, 2020. AAAI Press, pp 8783–8790, https://ojs.aaai.org/index.php/AAAI/article/view/6405
43. Kochkina E, Liakata M, Augenstein I (2017) Turing at semeval-2017 task 8: sequential approach to rumour stance classification with branch-lstm. CoRR arXiv:1704.07221
44. Kochkina E, Liakata M, Zubiaga A (2018) All-in-one: multi-task learning for rumour verification. CoRR arXiv:1806.03713
45. Kong SH, Tan LM, Gan KH, et al (2020) Fake news detection using deep learning. In: 2020 IEEE 10th symposium on computer applications & industrial electronics (ISCAIE), pp 102–107, https://doi.org/10.1109/ISCAIE47305.2020.9108841
46. Kotteti CMM, Dong X, Qian L (2018) Multiple time-series data analysis for rumor detection on social media. In: IEEE international conference on big data (IEEE BigData 2018), Seattle, WA, USA, December 10–13, 2018. IEEE, pp 4413–4419, https://doi.org/10.1109/BigData.2018.8622631
47. Kotteti CMM, Dong X, Qian L (2019) Rumor detection on time-series of tweets via deep learning. In: 2019 IEEE military communications conference, MILCOM 2019, Norfolk, VA, USA, November 12–14, 2019. IEEE, pp 1–7, https://doi.org/10.1109/MILCOM47813.2019.9020895
48. Kotteti CMM, Dong X, Qian L (2020) Ensemble deep learning on time-series representation of tweets for rumor detection in social media. CoRR arXiv:2004.12500
49. Kumar A, Sangwan SR, Nayyar A (2019) Rumour veracity detection on twitter using particle swarm optimized shallow classifiers. Multim Tools Appl 78(17):24,083-24,101. https://doi.org/10.1007/s11042-019-7398-6
50. Kumar A, Bhatia MPS, Sangwan SR (2022) Rumour detection using deep learning and filter-wrapper feature selection in benchmark twitter dataset. Multim Tools Appl 81(24):34,615-34,632. https://doi.org/10.1007/s11042-021-11340-x
51. Kumar S, Carley KM (2019) Tree lstms with convolution units to predict stance and rumor veracity in social media conversations. In: Proceedings of the 57th conference of the association for computational linguistics, ACL 2019, Florence, Italy, July 28–August 2, 2019, Volume 1: Long Papers. Association for Computational Linguistics, pp 5047–5058, https://doi.org/10.18653/v1/p19-1498
52. Lao A, Shi C, Yang Y (2021) Rumor detection with field of linear and non-linear propagation. In: WWW '21: the web conference 2021, Virtual Event/Ljubljana, Slovenia, April 19–23, 2021. ACM/IW3C2, pp 3178–3187, https://doi.org/10.1145/3442381.3450016
53. Li G, Dong M, Ming L et al (2022) Deep reinforcement learning based ensemble model for rumor tracking. Inf Syst 103(101):772. https://doi.org/10.1016/j.is.2021.101772
54. Li Q, Zhang Q, Si L (2019) Rumor detection by exploiting user credibility information, attention and multi-task learning. In: Proceedings of the 57th conference of the association for computational lin-

guistics, ACL 2019, Florence, Italy, July 28–August 2, 2019, Volume 1: Long Papers. Association for Computational Linguistics, pp 1173–1179, https://doi.org/10.18653/v1/p19-1113

55. Liu J, Wang C, Li C et al (2021) DTN: deep triple network for topic specific fake news detection. J Web Semant 70(100):646. https://doi.org/10.1016/j.websem.2021.100646
56. Liu X, Nourbakhsh A, Li Q, et al (2015) Real-time rumor debunking on twitter. In: Proceedings of the 24th ACM international conference on information and knowledge management, CIKM 2015, Melbourne, VIC, Australia, October 19–23, 2015. ACM, pp 1867–1870, https://doi.org/10.1145/2806416.2806651
57. Liu Y, Jin X, Shen H (2019) Towards early identification of online rumors based on long short-term memory networks. Inf Process Manag 56(4):1457–1467. https://doi.org/10.1016/j.ipm.2018.11.003
58. Lukasik M, Cohn T, Bontcheva K (2015) Classifying tweet level judgements of rumours in social media. In: Proceedings of the 2015 conference on empirical methods in natural language processing, EMNLP 2015, Lisbon, Portugal, September 17–21, 2015. The Association for Computational Linguistics, pp 2590–2595, https://doi.org/10.18653/v1/d15-1311
59. Lv Y, Sun X, Wen Y, et al (2022) Rumor detection based on time graph attention network. In: 2022 4th international conference on advances in computer technology, information science and communications (CTISC), pp 1–5, https://doi.org/10.1109/CTISC54888.2022.9849683
60. Ma J, Gao W (2020) Debunking rumors on twitter with tree transformer. In: Proceedings of the 28th international conference on computational linguistics, COLING 2020, Barcelona, Spain (Online), December 8–13, 2020. International Committee on Computational Linguistics, pp 5455–5466, https://doi.org/10.18653/v1/2020.coling-main.476
61. Ma J, Gao W, Mitra P, et al (2016) Detecting rumors from microblogs with recurrent neural networks. In: Proceedings of the twenty-fifth international joint conference on artificial intelligence, IJCAI 2016, New York, NY, USA, 9–15 July 2016. IJCAI/AAAI Press, pp 3818–3824, http://www.ijcai.org/Abstract/16/537
62. Ma J, Gao W, Wong K (2017) Detect rumors in microblog posts using propagation structure via kernel learning. In: Proceedings of the 55th annual meeting of the association for computational linguistics, ACL 2017, Vancouver, Canada, July 30–August 4, Volume 1: Long Papers. Association for computational linguistics, pp 708–717, https://doi.org/10.18653/v1/P17-1066
63. Ma J, Gao W, Wong K (2018a) Detect rumor and stance jointly by neural multi-task learning. In: Companion of the the web conference 2018 on the web conference 2018, WWW 2018, Lyon , France, April 23–27, 2018. ACM, pp 585–593, https://doi.org/10.1145/3184558.3188729
64. Ma J, Gao W, Wong K (2018b) Rumor detection on twitter with tree-structured recursive neural networks. In: Proceedings of the 56th annual meeting of the association for computational linguistics, ACL 2018, Melbourne, Australia, July 15–20, 2018, Volume 1: Long Papers. Association for Computational Linguistics, pp 1980–1989, https://aclanthology.org/P18-1184/
65. Madani Y, Erritali M, Bouikhalene B (2021) Using artificial intelligence techniques for detecting covid-19 epidemic fake news in moroccan tweets. Results Phys 25:104–266. https://www.sciencedirect.com/science/article/pii/S2211379721004034
66. Mikolov T, Chen K, Corrado G, et al (2013) Efficient estimation of word representations in vector space. In: 1st international conference on learning representations, ICLR 2013, Scottsdale, Arizona, USA, May 2–4, 2013, Workshop Track Proceedings, arXiv:1301.3781
67. Monti F, Frasca F, Eynard D, et al (2019) Fake news detection on social media using geometric deep learning. CoRR arXiv:1902.06673
68. 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):100,007. https://www.sciencedirect.com/science/article/pii/S2667096820300070
69. de Oliveira NR, Pisa PS, Lopez MA et al (2021) Identifying fake news on social networks based on natural language processing: trends and challenges. Inf 12(1):38. https://doi.org/10.3390/info12010038
70. Öztürk P, Li H, Sakamoto Y (2015) Combating rumor spread on social media: the effectiveness of refutation and warning. In: 48th Hawaii international conference on system sciences, HICSS 2015, Kauai, Hawaii, USA, January 5–8, 2015. IEEE Computer Society, pp 2406–2414, https://doi.org/10.1109/HICSS.2015.288
71. Pennington J, Socher R, Manning CD (2014) Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing, EMNLP 2014, October 25–29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL. ACL, pp 1532–1543, https://doi.org/10.3115/v1/d14-1162
72. Peters ME, Neumann M, Iyyer M, et al (2018) Deep contextualized word representations. In: Proceedings of the 2018 conference of the north american chapter of the association for computational linguistics: human language technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018,

Volume 1 (Long Papers). Association for Computational Linguistics, pp 2227–2237, https://doi.org/10.18653/v1/n18-1202

73. Potthast M, Kiesel J, Reinartz K, et al (2018) A stylometric inquiry into hyperpartisan and fake news. In: Proceedings of the 56th annual meeting of the association for computational linguistics, ACL 2018, Melbourne, Australia, July 15–20, 2018, Volume 1: Long Papers. Association for Computational Linguistics, pp 231–240, https://aclanthology.org/P18-1022/
74. Qazvinian V, Rosengren E, Radev DR, et al (2011) Rumor has it: Identifying misinformation in microblogs. In: Proceedings of the 2011 conference on empirical methods in natural language processing, EMNLP 2011, 27–31 July 2011, John McIntyre Conference Centre, Edinburgh, UK, A meeting of SIGDAT, a Special Interest Group of the ACL. ACL, pp 1589–1599, https://aclanthology.org/D11-1147/
75. Rahman MM, Watanobe Y, Nakamura K (2021) A bidirectional LSTM language model for code evaluation and repair. Symmetry 13(2):247. https://doi.org/10.3390/sym13020247
76. Rosenfeld N, Szanto A, Parkes DC (2020) A kernel of truth: determining rumor veracity on twitter by diffusion pattern alone. In: WWW '20: the web conference 2020, Taipei, Taiwan, April 20–24, 2020. ACM/IW3C2, pp 1018–1028, https://doi.org/10.1145/3366423.3380180
77. Roy S, Bhanu M, Saxena S et al (2022) gdart: improving rumor verification in social media with discrete attention representations. Inf Process Manag 59(3):102,927. https://doi.org/10.1016/j.ipm.2022.102927
78. Salem FKA, Feel RA, Elbassuoni S, et al (2019) FA-KES: A fake news dataset around the syrian war. In: Proceedings of the thirteenth international conference on web and social media, ICWSM 2019, Munich, Germany, June 11–14, 2019. AAAI Press, pp 573–582, https://ojs.aaai.org/index.php/ICWSM/article/view/3254
79. Shelke S, Attar V (2022) Rumor detection in social network based on user, content and lexical features. Multim Tools Appl 81(12):17,347-17,368. https://doi.org/10.1007/s11042-022-12761-y
80. Shim J, Lee Y, Ahn H (2021) A link2vec-based fake news detection model using web search results. Expert Syst Appl 184(115):491. https://doi.org/10.1016/j.eswa.2021.115491
81. Shu K, Sliva A, Wang S et al (2017) Fake news detection on social media: A data mining perspective. SIGKDD Explor 19(1):22–36. https://doi.org/10.1145/3137597.3137600
82. Shu K, Mahudeswaran D, Wang S et al (2020) Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big Data 8(3):171–188. https://doi.org/10.1089/big.2020.0062
83. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: 3rd international conference on learning representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings, arXiv:1409.1556
84. Song Y, Chen Y, Chang Y, et al (2021) Adversary-aware rumor detection. In: Findings of the association for computational linguistics: ACL/IJCNLP 2021, Online Event, August 1–6, 2021, Findings of ACL, vol ACL/IJCNLP 2021. Association for Computational Linguistics, pp 1371–1382, https://doi.org/10.18653/v1/2021.findings-acl.118
85. Su Q, Wan M, Liu X et al (2020) Motivations, methods and metrics of misinformation detection: an nlp perspective. Nat Lang Process Res 1:1–13. https://doi.org/10.2991/nlpr.d.200522.001
86. Sujana Y, Li J, Kao H (2020) Rumor detection on twitter using multiloss hierarchical bilstm with an attenuation factor. In: Proceedings of the 1st conference of the asia-pacific chapter of the association for computational linguistics and the 10th international joint conference on natural language processing, AACL/IJCNLP 2020, Suzhou, China, December 4-7, 2020. Association for Computational Linguistics, pp 18–26, https://aclanthology.org/2020.aacl-main.3/
87. Tan L, Wang G, Jia F, et al (2022) Research status of deep learning methods for rumor detection. CoRR arXiv:2204.11540. https://doi.org/10.48550/arXiv.2204.11540
88. Tripathy RM, Bagchi A, Mehta S (2010) A study of rumor control strategies on social networks. In: Proceedings of the 19th ACM conference on information and knowledge management, CIKM 2010, Toronto, Ontario, Canada, October 26–30, 2010. ACM, pp 1817–1820, https://doi.org/10.1145/1871437.1871737
89. Tripathy RM, Bagchi A, Mehta S (2013) Towards combating rumors in social networks: models and metrics. Intell Data Anal 17(1):149–175. https://doi.org/10.3233/IDA-120571
90. Tu K, Chen C, Hou C et al (2021) Rumor2vec: a rumor detection framework with joint text and propagation structure representation learning. Inf Sci 560:137–151. https://doi.org/10.1016/j.ins.2020.12.080
91. Varshney D, Vishwakarma DK (2021) A review on rumour prediction and veracity assessment in online social network. Expert Syst Appl 168(114):208. https://doi.org/10.1016/j.eswa.2020.114208
92. Veyseh APB, Ebrahimi J, Dou D, et al (2017) A temporal attentional model for rumor stance classification. In: Proceedings of the 2017 ACM on conference on information and knowledge management, CIKM

2017, Singapore, November 06–10, 2017. ACM, pp 2335–2338, https://doi.org/10.1145/3132847.3133116

93. Wang G, Tan L, Song T, et al (2022) Region-enhanced deep graph convolutional networks for rumor detection. CoRR arXiv:2206.07665. https://doi.org/10.48550/arXiv.2206.07665
94. Wang W, Qiu Y, Xuan S et al (2021) Early rumor detection based on deep recurrent q-learning. Secur Commun Netw 2021:5569,064:1-5569,064:13. https://doi.org/10.1155/2021/5569064
95. 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, ACL 2017, Vancouver, Canada, July 30–August 4, Volume 2: Short Papers. Association for Computational Linguistics, pp 422–426, https://doi.org/10.18653/v1/P17-2067
96. Wang Y, Wang L, Yang Y et al (2021) Semseq4fd: integrating global semantic relationship and local sequential order to enhance text representation for fake news detection. Expert Syst Appl 166(114):090. https://doi.org/10.1016/j.eswa.2020.114090
97. Wei L, Hu D, Zhou W, et al (2021) Towards propagation uncertainty: Edge-enhanced bayesian graph convolutional networks for rumor detection. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021. Association for Computational Linguistics, pp 3845–3854, https://doi.org/10.18653/v1/2021.acl-long.297
98. Wu L, Morstatter F, Carley KM et al (2019) Misinformation in social media: definition, manipulation, and detection. SIGKDD Explor 21(2):80–90. https://doi.org/10.1145/3373464.3373475
99. Wu L, Rao Y, Nazir A et al (2020) Discovering differential features: adversarial learning for information credibility evaluation. Inf Sci 516:453–473. https://doi.org/10.1016/j.ins.2019.12.040
100. Wu Z, Pi D, Chen J et al (2020) Rumor detection based on propagation graph neural network with attention mechanism. Expert Syst Appl 158(113):595. https://doi.org/10.1016/j.eswa.2020.113595
101. www.ETGovernment.com (2020) Rumors vs fake news: How to address misinformation in crisis? ETGovernment https://government.economictimes.indiatimes.com/news/digital-india/rumors-vs-fake-news-how-to-address-misinformation-in-crisis/76421449
102. Xuan K, Xia R (2019) Rumor stance classification via machine learning with text, user and propagation features. In: 2019 international conference on data mining workshops, ICDM Workshops 2019, Beijing, China, November 8-11, 2019. IEEE, pp 560–566, https://doi.org/10.1109/ICDMW.2019.00085
103. Yang F, Liu Y, Yu X, et al (2012) Automatic detection of rumor on sina weibo. In: Proceedings of the ACM SIGKDD workshop on mining data semantics. association for computing machinery, New York, NY, USA, MDS '12, https://doi.org/10.1145/2350190.2350203
104. Yang Y, Zheng L, Zhang J, et al (2018) TI-CNN: convolutional neural networks for fake news detection. CoRR arXiv:1806.00749
105. Yang Y, Wang Y, Wang L et al (2022) Postcom2dr: utilizing information from post and comments to detect rumors. Expert Syst Appl 189(116):071. https://doi.org/10.1016/j.eswa.2021.116071
106. Yu D, Zhou Y, Zhang S et al (2022) Heterogeneous graph convolutional network-based dynamic rumor detection on social media. CompLex 2022:8393,736:1-8393,736:10. https://doi.org/10.1155/2022/8393736
107. Zeng H, Cui X (2022) Simclrt: a simple framework for contrastive learning of rumor tracking. Eng Appl Artif Intell 110(104):757. https://doi.org/10.1016/j.engappai.2022.104757
108. Zeng J, Zhang Y, Ma X (2020) Fake news detection for epidemic emergencies via deep correlations between text and images. Sustainable Cities and Society p 102652
109. Zhang D, Nan F, Wei X, et al (2021) Supporting clustering with contrastive learning. In: Proceedings of the 2021 conference of the north american chapter of the association for computational linguistics: human language technologies, NAACL-HLT 2021, Online, June 6–11, 2021. Association for Computational Linguistics, pp 5419–5430, https://doi.org/10.18653/v1/2021.naacl-main.427
110. Zhang H, Fang Q, Qian S, et al (2019) Multi-modal knowledge-aware event memory network for social media rumor detection. In: Proceedings of the 27th ACM international conference on multimedia, MM 2019, Nice, France, October 21–25, 2019. ACM, pp 1942–1951, https://doi.org/10.1145/3343031.3350850
111. Zhang X, Zhao JJ, LeCun Y (2015) Character-level convolutional networks for text classification. CoRR arXiv:1509.01626
112. Zhao L, Yin J, Song Y (2016) An exploration of rumor combating behavior on social media in the context of social crises. Comput Hum Behav 58:25–36. https://doi.org/10.1016/j.chb.2015.11.054
113. Zhou C, Li K, Lu Y (2021) Linguistic characteristics and the dissemination of misinformation in social media: The moderating effect of information richness. Inf Process Manag 58(6):102,679. https://doi.org/10.1016/j.ipm.2021.102679

114. Zubiaga A, Hoi GWS, Liakata M, et al (2015) Analysing how people orient to and spread rumours in social media by looking at conversational threads. CoRR arXiv:1511.07487
115. Zubiaga A, Kochkina E, Liakata M, et al (2016a) Stance classification in rumours as a sequential task exploiting the tree structure of social media conversations. In: COLING 2016, 26th international conference on computational linguistics, proceedings of the conference: Technical Papers, December 11–16, 2016, Osaka, Japan. ACL, pp 2438–2448, https://aclanthology.org/C16-1230/
116. Zubiaga A, Liakata M, Procter R (2016b) Learning reporting dynamics during breaking news for rumour detection in social media. CoRR arXiv:1610.07363
117. Zubiaga A, Liakata M, Procter R (2017) Exploiting context for rumour detection in social media. In: Social informatics - 9th international conference, SocInfo 2017, Oxford, UK, September 13–15, 2017, Proceedings, Part I, Lecture Notes in Computer Science, vol 10539. Springer, pp 109–123, https://doi.org/10.1007/978-3-319-67217-5_8
118. Zubiaga A, Aker A, Bontcheva K et al (2018) Detection and resolution of rumours in social media: A survey. ACM Comput Surv 51(2):32:1-32:36. https://doi.org/10.1145/3161603

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

Barsha Pattanaik is a Ph.D. Scholar in Computer Science and Engineering at XIM University, Bhubaneswar, Odisha, India. She received M.Tech. degree in Electronics and Communication Engineering from Gandhi Institute of Engineering and Technology, Odisha, India, in 2010. Her research interest includes text classification, sentiment analysis in the field of Natural language processing, and Artificial intelligence.

Sourav Mandal has been an Assistant Professor at XIM University's School of Computer Science and Engineering (SCSE), in Bhubaneswar, Odisha, India since October 2020. Prior to that, he had been employed since 2006 as an Assistant Professor in the Department of Computer Science and Engineering at the Haldia Institute of Technology in Haldia, India. Among his research interests in the natural language processing (NLP) and artificial intelligence (AI) field are natural language understanding, information extraction, text classification, text summarization, etc. with data science, machine learning, and deep learning. Sourav Mandal earned a bachelor's degree in Computer Science & Engineering from The University of Burdwan in Burdwan, India, in 2003, a master's degree in Multimedia Development from Jadavpur University in Kolkata, India, in 2005, and a Ph.D. in engineering from Jadavpur University in Kolkata, India, in 2020.

Rudra M. Tripathy holds the positions of Associate Professor and Dean (Academic) in the School of Computer Science and Engineering, XIM University, Bhubaneswar. He has more than 20 years of teaching experiences in various reputed institutes. He served as Head of the Department of Computer Science and Engineering, Silicon Institute of Technology for more than 6 years. He holds a Ph.D. in Computer Science and Engineering from I I T Delhi. His research focuses on the area of Data Mining, Machine Learning, Social Network Analysis and Structural Properties of Networks. His research work on "Rumor Control Strategies" has been covered by many news media: The Hindu, The Indian Express, BBS Fake News. He has published many papers in reputed international conferences and Journals.