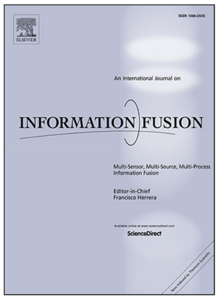
Journal Pre-proof



Emotion detection for misinformation: A review

Zhiwei Liu, Tianlin Zhang, Kailai Yang, Paul Thompson, Zeping Yu,
Sophia Ananiadou

PII: S1566-2535(24)00078-2
DOI: https://doi.org/10.1016/j.inffus.2024.102300
Reference: INFFUS 102300

To appear in: *Information Fusion*

Received date : 3 November 2023
Revised date : 7 February 2024
Accepted date : 9 February 2024

Please cite this article as: Z. Liu, T. Zhang, K. Yang et al., Emotion detection for misinformation: A review, *Information Fusion* (2024), doi: https://doi.org/10.1016/j.inffus.2024.102300.

Emotion Detection for Misinformation: A Review

Zhiwei Liu[a], Tianlin Zhang[a], Kailai Yang[a], Paul Thompson[a], Zeping Yu[a] and Sophia Ananiadou[a,*]

[a]National Centre for Text Mining, Department of Computer Science, The University of Manchester, Manchester, M1 7DN, UK

ARTICLE INFO

Keywords:
Sentiment analysis
Emotion detection
Misinformation
Rumor
Fake news
Stance detection

ABSTRACT

With the advent of social media, an increasing number of netizens are sharing and reading posts and news online. However, the huge volumes of misinformation (e.g., fake news and rumors) that flood the internet can adversely affect people's lives, and have resulted in the emergence of rumor and fake news detection as a hot research topic. The emotions and sentiments of netizens, as expressed in social media posts and news, constitute important factors that can help to distinguish fake news from genuine news and to understand the spread of rumors. This article comprehensively reviews emotion-based methods for misinformation detection, with a particular focus on advanced fusion methods. We begin by explaining the strong links between emotions and misinformation. We subsequently provide a detailed analysis of a range of misinformation detection methods that employ a variety of emotion, sentiment and stance-based features, and describe their strengths and weaknesses. Finally, we discuss a number of ongoing challenges in emotion-based misinformation detection based on large language models, and suggest future research directions, including data collection (multi-platform, multilingual), annotation, benchmark, multimodality, and interpretability.

1. Introduction

Misinformation is false information that is created specifically to mislead readers [1], including *fake news* and *rumors*. Fake news refers to intentionally fabricated information whose publishing or dissemination may mislead readers or result in panic [2]. Rumors are defined as unverified or unsupported hearsay or information that become spread among people [3]. Rumors and fake news are now ubiquitous. They affect people's daily lives, alter their emotions, and lead them to trust incorrect information. Social media platforms, such as Twitter, Facebook, Reddit, and Sina Weibo, constitute important means not only for socializing, but also for spreading news and rumors, and generate a huge amount of information every day [4]. According to the Datareportal April 2023 global overview[1], approximately 4.80 billion people (about 60% of the world's population) use social media. Moreover, its use is continuing to grow rapidly, with 150 million new user identities added in the last year, representing an annual growth rate of 3.2%. Now that smartphones are commonplace, users can create, share and browse publicly available content on social media anytime and anywhere, thus increasing the ease and speed at which information can spread. However, due to a lack of effective regulatory measures, the Internet has become flooded with fake news and rumors, which can be challenging to distinguish from genuine facts [5]. Such misinformation can manipulate the emotions and intentions of netizens [6], which in turn can impact upon social factors, politics and the economy. For example, during the COVID-19 pandemic, rumors about the virus spread across the Internet, which caused panic and tension among society [7]. Furthermore, recent advances in artificial intelligence (AI) and the emergence of large language models (LLMs), such as Instruct-GPT [8], ChatGPT, and GPT-4[9], are making it increasingly straightforward to generate false information that appears highly convincing [10]. Accordingly, there is an urgent global-level need for methods that can detect misinformation effectively.

Rumors and fake news trigger specific emotions and sentiments. For example, Zaeem et al. [11] observe a statistically significant relationship between negative sentiment and fake news. These emotions and sentiments can in turn give rise to specific behaviors or actions, such as the motivation to spread rumors [12]. Furthermore, readers are more likely to believe news that aligns with their existing beliefs [13]. For instance, in politics, conservative supporters are more likely to believe negative news about liberals. Rumor-mongering often takes advantage of these trends by disseminating fake news on social media channels that targets users with particular beliefs, and which triggers strong emotions [6]. For example, fake news that attacks politics often intentionally embeds anger [14]. The aim of the rumor-mongers is to promote the further spread of the rumor by encouraging user actions such as forwarding, liking, and commenting. This behavior is exemplified in Figure 1, which shows two samples of fake news on social media, with associated user comments. It has been found that false rumors tend to generate more reshares, spread over longer time periods, and become more viral, when they include words that convey emotions of trust, anticipation, or anger [15, 16]. Additionally, it was found that during the COVID-19 epidemic, there was a correlation between the level of anger felt by the public and the likelihood that rumors would be circulated [17]. All of the above observations serve to demonstrate the strong relationships between emotions and misinformation.

Recently, it has been shown that natural language processing (NLP) methods that recognize *affective* information

*Corresponding author
sophia.ananiadou@manchester.ac.uk (S. Ananiadou)
https://research.manchester.ac.uk/en/persons/sophia.ananiadou (S. Ananiadou)
ORCID(s):

[1]https://datareportal.com/reports/digital-2023-april-global-statshot

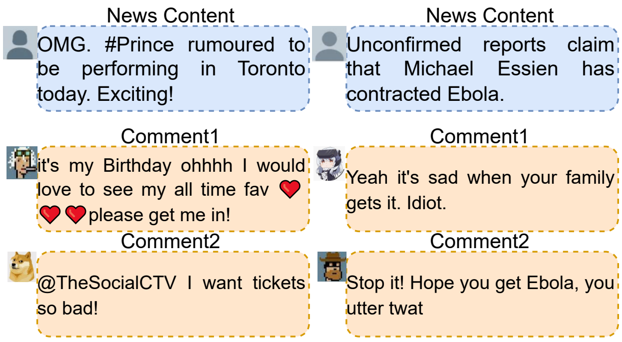


Figure 1: Fake news samples

(e.g., emotions and sentiment [18]) in text can make important contributions towards the automated detection of misinformation and conspiracies [19]. Significant advances in many NLP tasks (e.g., classification, summarisation, question answering, and information extraction) have been facilitated by the advent of deep learning (DL) methods, which are able to extract higher-level and more complex feature representations through multiple processing layers, compared to conventional machine learning (ML) methods. Various DL approaches that exploit emotion features have been used to approach the problem of misinformation detection, including Convolutional Neural Networks (CNN) [20], Recurrent Neural Networks (RNN) [21], and Graph Convolutional Networks (GCN) [22]. Furthermore, pre-trained language models such as BERT[23], RoBERTa [24], and LLMs [25, 26, 27] have been used as backbone models for detecting misinformation. The various proposed methods exploit emotion features in diverse ways. For example, Al-Saif et al. [28] describe a context-aware approach for rumor detection in Arabic social media that combines emotion features with other types of features (i.e., topics and reactions), while Zhang et al. [29] account for the *dual emotions* expressed in both a fake news post and its follow-up comments. Emotion detection can also be successfully employed as an auxiliary task within a multitask framework to improve the accuracy of fake news detection [30]. Such examples illustrate the potential for emotion information to be integrated within misinformation detection methods in a broad range of ways to improve performance.

Determining the *stance* of social media users towards news also plays a crucial role in identifying misinformation [31]. Stance is defined as the expression of an attitude towards a given piece of information [32], which may include supporting, denying, querying, or commenting upon it [33]. Users often take a stance towards rumors propagated in online spheres [34]. For example, the public has expressed various attitudes towards climate change on social media platforms [35]. Moreover, users are more likely to accept and support information that aligns with their viewpoints [36]. For instance, individuals with strong opinions about "Americanness" tended to demonstrate support in their tweets relating to former US President Trump's October 2018 post concerning the cancellation of birthright citizenship [37]. Emotions and sentiment have an underlying connection with attitudes, and are thus advantageous for stance detection [38, 39]. For example, if a person expresses positive feelings towards a political candidate, then this is likely to indicate that they support or agree with the candidate's policies. The importance of sentiment and emotion has been confirmed by a number of studies that have used them in combination with other features for stance detection in rumors and fake news[40, 41, 42]. Examples include Wang et al. [43], who combine emotion and sentiment with Twitter metadata features; Xuan et al. [44], who integrate emotion with content and user features; and Parimi et al. [45], who make use of various features of rumors, including content and emotions.

1.1. Comparisons with Previous Surveys

Several surveys relating to rumor and fake news detection have been published recently, most of which review and summarize a variety of detection techniques from a macro perspective, i.e., they provide general overviews of the various models and/or features that have been employed for misinformation detection [46, 47, 48, 49, 50]. Although some of these studies mention emotion information among other text-based features, they do not specifically focus on the importance of emotion for misinformation detection.

Other reviews have a more specific focus. For example, Varlamis et al. [51] cover fake news detection methods that employ GCN and graph information, while D'Ulizia et al. [52] provide an overview of available datasets for evaluating fake news detection methods. Alsaif et al. [53] review recent approaches that use stance detection as a means to identify rumors, while Hardalov et al. [54] examine the relationship between stance detection and misinformation identification. Meanwhile, Shahid et al. [55] conduct a comprehensive survey of state-of-the-art methods for detecting malicious users and bots, and Shelke et al. [56] analyze methods for detecting sources of misinformation in social networks. Among these more specific studies, little attention is paid to the role of emotion in fake news and rumor detection.

To our knowledge, the only previous survey that specifically investigates and highlights the importance of affective features for misinformation detection is conducted by Alonso et al. [19], who provide an overview of the use of sentiment in the detection of fake news. However, as explained in more detail in Section 2.1, sentiments (generally corresponding to positive, negative or neutral tones) are more coarse-grained than emotions, which consist of a more fine-grained range of specific and stronger feelings. In [19], approaches that utilize emotions for misinformation detection are only very briefly touched upon, while emotion-based stance detection is not discussed. Furthermore, given the highly active nature of research in this area, there are many recently developed methods that are not included in this previous survey.

Accordingly, there is currently a lack of an existing survey that comprehensively explores methods that use *both* types of affective information (i.e., sentiment and fine-grained emotions) to guide the detection of misinformation, and which also explores the role of emotions in stance detection.

1.2. Research Questions

Based on the research gaps of existing reviews identified in the previous section, we sought to carry out a survey that would delve more deeply into the importance of both emotion and sentiment in the detection of misinformation, and that would comprehensively explore methods that have made use of these types of affective information in different ways. To guide our survey, we posit the following research questions (RQs):

RQ 1: Which types of relationships exist between sentiment/ emotion and fake news/rumors?

RQ 2: Which existing methods for misinformation detection are emotion-based? How can such methods be categorized? What are the differences between these methods?

RQ 3: What are the ongoing challenges and future directions of emotion-based misinformation detection?

RQ 4: How can LLMs contribute to misinformation detection?

To address the above RQs, we have collected and analyzed relevant articles from a number of different literature search platforms. To address RQ 1, we initially analyze articles that explore the relationships between affective information and misinformation. To answer RQ 2, we subsequently summarize and categorize different emotion-based misinformation detection methods (with an emphasis on advanced fusion techniques), report and compare their performance, and analyze their relative strengths and weaknesses. Finally, to respond to RQ 3 and RQ 4, we discuss the current challenges faced in misinformation detection and identify important future directions of research, with a particular focus on how LLMs can play a part in helping to advance the field.

1.3. Literature Collection Strategies

We collected articles from five different literature search platforms, i.e., IEEE Xplore, ACM Digital Library, Web of Science, Scopus, and DBLP. The article selection process consisted of three main steps, i.e.: *collection, preliminary screening*, and *manual review*.

Collection: Similarly to the search strategy described in [57], we conducted an initial keyword search aimed at retrieving articles published between January 2016 and September 2023 that mention both misinformation and affective information. The specific query used was as follows: *(emotion OR sentiment OR affective) AND (rumor OR "fake news" OR misinformation OR disinformation)*. The search resulted in the retrieval of 6,483 articles (326 from IEEE Xplore; 3345 from ACM Digital Library; 1036 from Web of Science; 1698 from Scopus; and 78 from DBLP.)

Preliminary screening: After deduplication, we employed RobotAnalyst [58], a tool that prioritizes articles based on relevance feedback and active learning [59, 60] in order to minimize the amount of human work required in the screening phase of reviews. Articles were screened based on title and abstract, and were retained only if: (1) They were relevant to rumor/fake news analysis or detection. (2) They involved the use of affective information. The screening process resulted in the identification of 473 articles for further review.

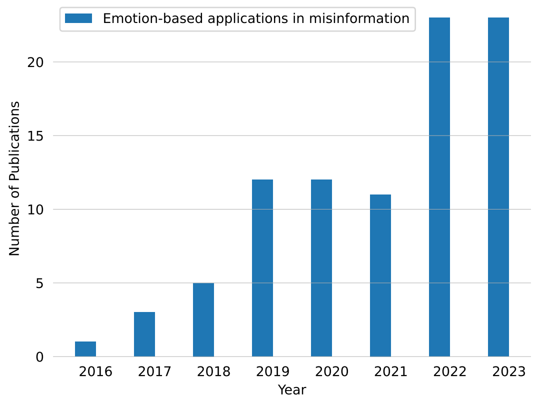


Figure 2: Distribution of publications on emotion-based applications in misinformation published since 2016.

Manual review: We conducted a manual full-text examination of the articles resulting from the preliminary screening phase, and retained those that: (1) Focus on methods both for analyzing and/or detecting misinformation, *and* for detecting emotions and/or sentiment. (2) Apply learning methods to the task of misinformation detection or analysis. (3) Use affective information as a feature for misinformation detection or analysis. By applying these criteria, 90 articles were retained, and form the basis for the detailed analysis presented in this review.

Figure 2 illustrates the temporal distribution of studies describing emotion-based applications in misinformation that have been published in recent years. Particularly noticeable is the significant surge in the number of articles published over the last two years. This provides evidence of the increasing appreciation of the importance of emotion in detecting rumors and fake news.

1.4. Contributions of the Survey

This is the first comprehensive survey of methods that focuses on the use of different types of affective features (i.e., both emotion and sentiment) as a means to detect fake news, rumors, and stance, with an emphasis on advanced emotion-based fusion methods. The aim of the survey is to facilitate an enhanced understanding of the latest developments in this area and to act as a driver and a guide for future research endeavors. Our main contributions are as follows:

1. We summarize the findings of articles exploring relationships between affective information and misinformation,

which identify significant links between them. The importance of developing emotion-based methods for misinformation detection is thus confirmed.

2. We categorize and provide a detailed overview of the characteristics of available datasets that can support misinformation detection, including those that cover languages other than English or that include multimodal data. The majority of these datasets are publicly available, which offers considerable benefits in promoting research endeavors focused on misinformation detection.

3. We categorize and describe emotion-based methods for misinformation detection based on both conventional ML and DL methods, with a particular focus on advanced fusion approaches. We also provide an overview of articles concerning emotion-based stance detection in misinformation. We employ a combination of textual descriptions, tables, and images to describe the different methodologies, features, and datasets used.

4. We present and analyze the performance of the various emotion-based methods, and discuss their relative strengths and weaknesses. We compare the performance levels achieved by advanced fusion methods with those attained by methods based on conventional ML, DL, emotion-based stance detection and LLMs. We discover that advanced fusion methods exhibit the highest levels of performance and robustness, according to their combination of multiple relevant features and/or through their integration of various advanced techniques.

5. We outline a number of challenges faced in the development of misinformation detection methods, and suggest promising future research directions, including dataset collection (multi-platform, multilingual), emotion annotation, multimodality, benchmark construction, and interoperability. We take into account the increasingly important role of LLMs, and how to make the best of them. It is hoped that this discussion will help guide researchers to address the most pertinent issues when carrying out future research in the field of misinformation detection.

Organization: Figure 3 illustrates the structure of the remainder of the article, which may be summarised as follows: Section 2 introduces related work on the detection of emotions, sentiment, rumors, and fake news. Section 3 presents and analyzes studies that explore relationships between emotion/sentiment and misinformation. Section 4 provides a detailed exploration of approaches to emotion-based misinformation detection, beginning with a summary of available datasets, followed by an account of the different types of methods that have been developed, including a detailed analysis of advanced fusion methods, and a summary of emotion-based stance detection in misinformation. This section proceeds by presenting and comparing the performance of different types of methods, and discusses the relative strengths and weaknesses of various advanced fusion methods. Section 5 presents ongoing challenges and future research directions; Section 6 introduces several potential factors that could impact upon the validity of the comparative analysis carried out in Section 4, and explains how these factors can help to guide more effective future research; Section 7 concludes the article by summarizing our findings. The appendices provide details of the abbreviations and notations used in the article (Appendix A); specific types of content-based features used by different methods (Appendix B); an inventory of commonly used sentiment and emotion detection tools (Appendix C); and a description of the various metrics that have been used to evaluate misinformation and stance detection methods (Appendix D).

2. Related Work

2.1. Sentiments and Emotions

Sentiments and emotions are important and fundamental aspects of our lives. What we do and say reflects our emotions in some way. Emotion detection (ED) and sentiment analysis (SA) are two types of NLP techniques for analyzing human expressions that can help us to understand people's feelings towards specific topics [61]. SA [62] aims to capture the overall emotional tone conveyed by a data source (usually *positive*, *negative*, or *neutral*), along with the strength of this tone [63]. ED is the process of classifying data at a finer-grained level, according to the emotions that it conveys. Compared to sentiment, the term *emotion* refers to more specific and stronger feelings [64]. For example, positive sentiment encompasses a range of different emotions, such as *happiness* and *joy*, while negative sentiment includes the emotions of *sadness* and *anger*, among others.

A number of theoretical emotion classification models have been proposed, which can be divided into two categories, i.e., *categorical* and *dimensional* [65]. Categorical models define a single discrete set of emotional states; examples include Shaver [66] (sadness, love, joy, anger, surprise, and fear), Ekman [67] (joy, anger, fear, disgust, sadness, and surprise), and Plutchik [68] (anticipation, surprise, anger, fear, trust, disgust, joy, and sadness). In contrast, dimensional models posit that emotions can be decomposed into a number of distinct dimensions. One of the best-known examples is Plutchick's *wheel of emotions* [68, 69], in which emotions are defined within a two-dimensional space of *valence* and *arousal*. The wheel is divided into 24 primary, secondary, and tertiary dyads, based on eight basic emotions. Other popular dimensional emotion models include the *Pleasure-Arousal-Dominance (PAD)* model [70], which is based on three dimensions, i.e., *Pleasure* (the pleasantness of the emotion), *Arousal* (the level of physiological activation or intensity of the emotion), and *Domination* (the degree of control or dominance experienced through the emotion); and the *Valence-Arousal-Dominance (VAD)* model [71], in which *Arousal* and *Dominance* are supplemented by *Valence* (the positivity or negativity of the emotion).

A range of automated methods has been developed to detect both sentiment and emotions in text, which may be broadly categorized into dictionary-based, conventional ML [72], and DL [73] methods. The dictionary-based approach involves constructing an inventory of words that denote specific sentiments and/or emotions, and matching them

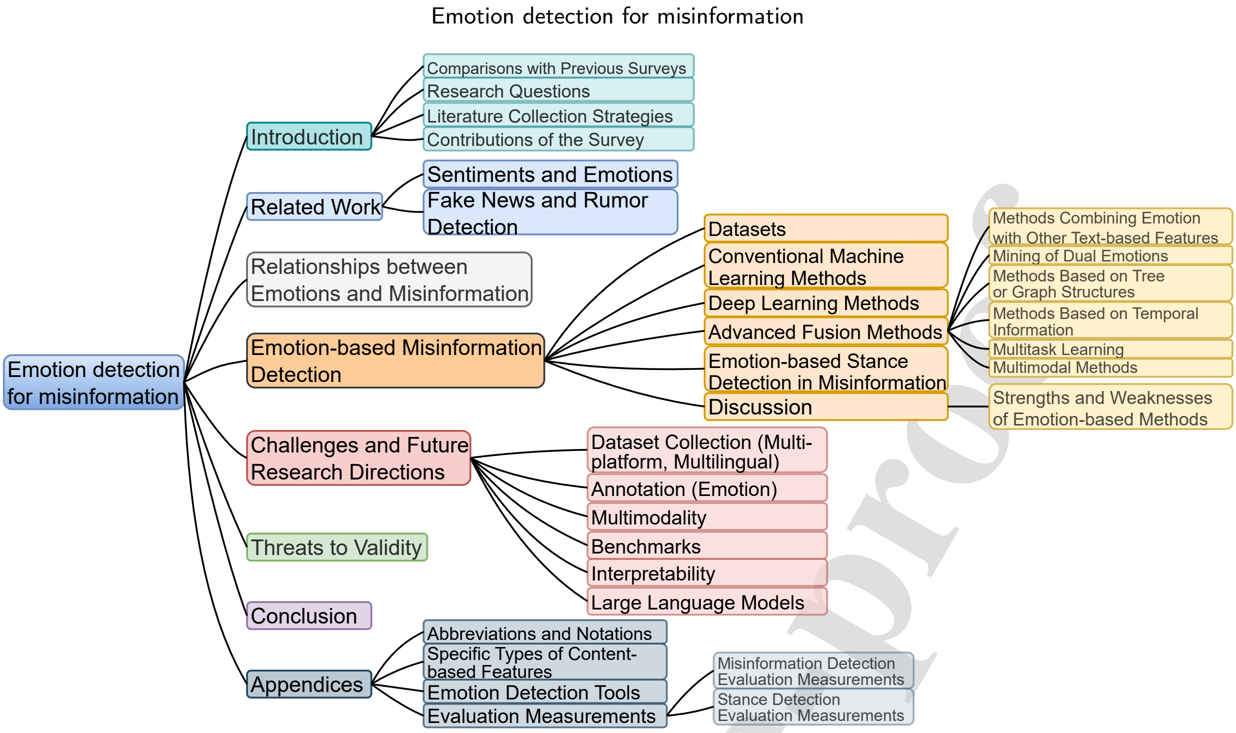


Figure 3: Structure of the article

against words appearing in the text to be processed to obtain information about the sentiments and emotions conveyed. Meanwhile, methods based on conventional ML and DL apply learning algorithms to datasets annotated with sentiment or emotion labels to teach them how to detect the different ways in which these types of emotions may be conveyed in text. Recently, there has also been a growing interest in exploring how LLMs can be exploited to enhance the accuracy of SA and ED [74, 75, 76].

2.2. Fake News and Rumor Detection

The convenience of accessing social media platforms on various electronic devices means that people can easily post or access large amounts of information on the Internet. This can lead to the rampant spread of misinformation. Certain individuals intentionally spread rumors to gain attention, mislead readers, or make a profit, even though such rumors can pose significant harm to society [77]. Therefore, there is an urgent need to detect misinformation in an efficient and effective manner. A large body of research has aimed to respond to this need, which has been summarized in various reviews. Most of these provide high-level overviews of methods, techniques and/or features that have been used to detect rumors and fake news [46, 47, 48, 49, 50, 51], while others discuss potential applications of these methods, including stance detection [53, 54] and source detection [55, 56]. The important role of affective information in misinformation detection is discussed in [19]. However, the review is almost exclusively focused on reviewing methods that utilize sentiment information to aid with misinformation detection, and covers few methods that employ more fine-grained emotion information as the basis for identifying fake news and/or rumors.

Misinformation detection approaches consist of three main components, i.e., the *datasets* used to support their development, the *methods* used to perform detection, and the *features* used within these methods. The majority of the datasets are obtained from social media platforms such as Twitter, Facebook, and Sina Weibo, or from fact-checking websites, such as Snopes[2], Factcheck[3], and PolitiFact[4]. Detection methods may be divided into those based on conventional ML [49] or DL [48, 51]. For instance, Kaliyar et al. propose a series of DL models (e.g. DeepFakE [78], AENeT [79], and FNDNet [80]), which make effective use of a range of misinformation features to improve detection accuracy. Figure 4 presents a range of features that have been employed for misinformation detection by different methods. Among these features, content-based features constitute the most diverse class; Table 7 in Appendix B provides further details regarding the specific types of features that fall under each of the content-based sub-classes shown in Figure 4. Within the *Affective* group of features, *dual emotion* features aim to account for the importance of considering different emotional perspectives in identifying misinformation, i.e., both *publisher emotion*, which refers to the emotions conveyed in an original post that starts a thread on social media, and *social emotion*, which refers to the emotions expressed

[2] https://www.snopes.com/
[3] https://www.factcheck.org/
[4] https://www.politifact.com/

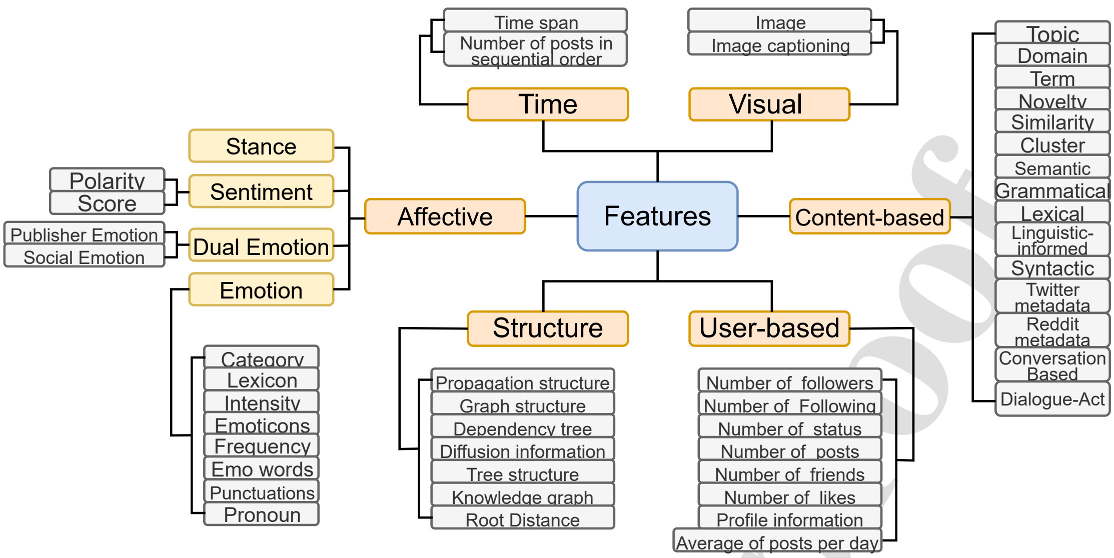


Figure 4: Features used in rumor and fake news detection

in follow-up posts that respond to and/or comment on the original post.

3. Relationships between Emotions and Misinformation

Although emotions are regarded as a dominant driver of human behavior, exploring their role in the online diffusion of misinformation has only recently begun. Misinformation can evoke emotional responses in readers, which in turn can lead to specific behaviors, such as belief in the information, resharing or liking it, etc. [13].

Table 1 lists a range of recent studies that have investigated the relationships between affective information and misinformation, e.g., how the expression of particular emotions or sentiments can indicate that a data source is likely to contain misinformation and/or predict the likely response of readers. For each study, we list the dataset used, the ED/SA and relationship analysis methods (RAM) employed, and details of the most important relationships identified. The most commonly explored topic is COVID-19, according to the explosion of rumors and fake news generated by the pandemic. To perform ED/SA, the majority of researchers apply dictionary-based methods (see Table 8 for details) or conventional ML methods, while Wu et al. [81] manually annotate discrete emotions based on the PAD emotional state model. In [12, 13, 82, 83, 84, 85], questionnaires are designed to ask participants to directly report their emotions. Among these approaches, Zhang et al. [12] and Martel et al. [82] use the Positive-Negative Emotional Scale (PANAS) to further quantify the emotional state of participants. The analysis of Li et al. [86] is based on the results obtained from their novel Multi-EmoBERT multi-label emotion recognition tool. Wan et al. [87] use a mixture of existing NLP tools and ED lexicons, enhanced using rules and automated weighting. They extract *Emotion Triple Elements* to study potentially different responses to emotional triggers. For relationship analysis, a range of commonly used statistical analysis methods are applied, including Logistic Regression (LR) [81, 88], Linear Regression [89], and the T-Test [90, 91].

Various indicators have been used to judge the impact of emotions on the spread of rumors or the degree of outbreak, etc. For example, the questionnaires of [12, 83, 85] directly ask participants which actions they would take when faced with certain types of news, such as sharing intentions or "likes". Other studies use cascade size, cascade lifetime, and structural virality [15, 16, 92] to analyze the patterns of misinformation spread. Cascade size corresponds to the number of forwardings generated by a cascade; cascade lifetime is the length of time that a rumor cascade remains active, i.e., the time elapsed between the root broadcast and the final forwarding; and structural virality [93] provides an aggregated metric combining the depth and breadth of a cascade. In addition, many studies analyze relationships by investigating the number of rumors that occur over time, or by comparing the number of rumors that convey different emotions.

The analyses detailed in Table 1 reveal a number of important relationships between affective information and misinformation, which can sometimes depend on the types of topics being discussed. Misinformation is generally associated with a significant level of high-arousal emotions such as anger, sadness, anxiety, surprise, and fear. Rumors conveying anger, sadness, anxiety, and fear are likely to generate a large number of shares, and to be long-lived and viral [15, 17, 81, 88], while emotional appeals (like anger and disgust) can increase users' engagement with fake posts [94]. Fake news expresses higher overall emotion, negative sentiment, and lower positive sentiment than genuine news

Table 1
Relationships between emotions and misinformation. ED/SA: Emotion Detection/Sentiment Analysis method, RAM: Relationship analysis method. MANOVA: Multivariate Analysis of Variance, MANCOVA: Multivariate Analysis of Covariance, ANOVA: Analysis of Variance. ANCOVA: Analysis of covariance.

Pub	Year	Data	ED/SA	RAM	Relationship (Partly)
[88]	2019	Demonetization related	LIWC	Logistic Regression	Posts with a higher level of anger, sadness, and anxiety are indicative of rumor.
[17]	2020	COVID-19 Related	Manual	Time-lagged Cross-correlation Analyses	The angrier, sadder, or more fear the public feels, the more rumors there are likely to be.
[82]	2020	News Headlines	Questionnaire,PANAS	Linear Mixed-effects Analyses	Emotion plays a causal role in people's susceptibility to incorrectly perceiving fake news as accurate.
[95]	2020	[96]	EmoLex	SVM	Emotion-based features contribute more to rumor recognition capabilities than personality-based ones.
[11]	2020	Open-Source Data	Meaningt.loud, TextBlob, AFINN	Chi-square Test, P(T|S), Goodman and Kruskal's Gamma	Relationships exist between negative sentiment and fake news, and between positive sentiment and genuine news.
[15]	2021	Twitter	Questionnaire	Generalized Linear Model	Rumors conveying anticipation, anger, or trust, or which are highly offensive, generate more shares, are longer-lived, and more viral.
[16]	2021	Twitter	EmoLex	Generalized Linear Model	False rumors with a high proportion of terms conveying positive sentiment, trust, anticipation, or anger are more likely to go viral.
[97]	2021	COVID-19 Related	Decision Tree	SPSS 22.0, Granger Causality Test	The more negative people feel about COVID-19, the more likely it is that rumors will be generated.
[13]	2021	News Headlines	Questionnaire	MANOVA, MANCOVA, ANOVA	Emotional reactivity of participants is associated with response behavior intentions.
[12]	2022	Questionnaire	Questionnaire, PANAS	Multilevel Linear Regression	Expression of emotion in online rumors positively affects readers' emotions. Readers' emotions affect their intentions to spread rumors.
[81]	2022	COVID-19 Related	Pleasure-Arousal-Dominance	Logistic Regression	Weibo messages filled with high-arousal emotions such as fear, anger and surprise are more likely to be rumors.
[89]	2022	Twitter, Weibo	Emotion Lexicon, ML, DL	Logistic Regression, Linear Regression	The ease with which fake news is spread online is positively associated with the strength of anger that it conveys.
[98]	2022	Open-Source Data	Textblob	Histogram	The negativity score of fake news is slightly higher than that of real news.
[92]	2022	COVID-19 Related	EmoLex	Generalized Linear Model	False rumors that include a large number of emotion words condemning others are more viral.
[85]	2022	Anti-vaccination	Questionnaire, LIWC	ANCOVA, SPSS24	Individuals who are more neutral towards vaccines and are angry are more likely to believe and share anti-vaccine fake news, compared to individuals who have anti-vaccine attitudes and are fearful.
[83]	2023	Political	Questionnaire	SPSS29, Structural Equation Modelling	Negative beliefs about the political system increase emotional responses to both genuine and fake news.
[90]	2023	COVID-19 Related	TextBlob,cn-sentiment-measures, LIWC	T-Test	Fake news expresses a higher level of overall emotion, negative emotion, and anger than genuine news.
[84]	2023	COVID-19 Related	Questionnaire	Partial Least Squares, Multigroup Analysis	Surprise is felt most intensely towards celebrity fake news and the toilet paper shortage rumor.
[86]	2023	Open-Source Data	Multi-EmoBERT Fig. 7	Chi-square Test	Fake news is associated with negative emotions and co-existing emotions in certain contexts.
[87]	2023	COVID-19 Related	Emotion Triple Elements	Inter-rater Reliability Analysis, Cohen's Kappa Coefficient	The emotion of fear plays an important role in the spread of fake news.
[99]	2023	SouthAfrican Website	VADER, T5 transformer	Word Clouds, T-SNE Plots and Histograms	Fake news in South Africa conveys more anger, joy, sadness, and fear than genuine news.
[91]	2023	COVID-19 Related	TextBlob, cn-sentiment-measures	T-Test	Fake news contains higher overall emotion, negative emotion, and anger than genuine news.
[94]	2023	COVID-19 Related	IBM Watson's NLU	Chi-square Test, Linear Regression	Anger and disgust increase users' engagement with fake posts.

[90, 91]. In general, it may be concluded that sentiment and emotions are both inextricably intertwined with misinformation, thus confirming their important roles in the automated detection of fake news and rumors.

4. Emotion-based Misinformation Detection

Motivated by the results of analyses such as those outlined in Section 3, many studies have used sentiment and/or emotions as the main features to guide the automated detection of fake information. In this section, we conduct a detailed survey of emotion-based methods for misinformation detection. We firstly introduce the datasets used to support the development of such methods, and subsequently describe a range of detection methods employing a mixture of conventional ML, DL, and advanced fusion techniques. We additionally provide a summary of the closely related task of emotion-based stance detection in misinformation. Table 3 lists the complete set of emotion-based misinformation detection methods that we have reviewed. Appendix C lists details of commonly used supporting ED and SA tools, while Appendix D provides an overview of the metrics used to evaluate misinformation detection methods.

4.1. Datasets

Table 2 lists a range of publicly available datasets aimed at facilitating the development and/or evaluation of misinformation detection methods. The majority of these datasets consist of data obtained from popular social media platforms and fact-checking websites, such as Twitter, Weibo, Reddit, politifact.com, gossipcop.com, etc. For each dataset, we provide its commonly used name and reference, its source, a description of its size and main characteristics, its level of availability, and notes. The latter are used to indicate datasets that cover languages other than English, those that are multimodal, those specifically concerning COVID-19, and those annotated with stance information. Datasets without notes consist of textual English data items that are labeled according to whether or not they represent misinformation.

As may be observed in Table 2, the majority of datasets are publicly available, which is highly advantageous to promote research in the field of misinformation. Due to the prevalence of rumors relating to the COVID-19 pandemic on social media, there has been a trend towards collecting misinformation datasets relating to this topic, as a means to explore rumor detection in the field of health disease transmission [100, 101, 102, 103]. Another important feature of several of the datasets listed (including PHEME [104], the Twitter series [105], and the Weibo series [29, 106, 107]) is their inclusion of comments/replies relating to original news stories or tweets. Such datasets allow the exploration of methods that take into account dual publisher and social emotions, and possible interactions between them, to improve the accuracy of misinformation detection. We can furthermore observe that several datasets are multimodal, i.e., they consist of both text and images. These include FakeNewsNet [108], Fakeddit [109], and MediaEval2016 [110]. The inclusion of images in these datasets provides scope to explore methods that take advantage of visual clues to complement text-based information in identifying misinformation. Although the majority of datasets contain only English text, there are a growing number of corpora that cover either single or multiple other languages, including Chinese, Portuguese, Spanish, and Danish, thus providing opportunities to develop methods that are multilingual and/or which target lesser resourced languages. While most datasets are annotated according to whether or not their constituent data items correspond to fake news or rumor, there are also a number of corpora annotated with stance-related labels, which can facilitate investigations into how stance information can contribute towards the detection of misinformation.

4.2. Conventional Machine Learning Methods

ML is a branch of AI that uses algorithms and statistical models to teach computers how to make predictions and decisions automatically. As shown in Table 3, a variety of conventional ML algorithms has been used to develop misinformation classifiers. These include both supervised methods, such as passive-aggressive [172], Naive Bayes (NB) [152], k-Nearest Neighbour (KNN) [136, 162], Support Vector Machine (SVM) [160], Random Forest (RF) [24], Decision Tree (DT) [158], AdaBoost (AB) [125], LR, XG-Boost, Gradient Boost (GB) [153], and unsupervised methods like K-Means and DBSCAN [126].

4.3. Deep Learning Methods

DL is a sub-field of ML that has made breakthrough progress in many fields, especially in computer vision, NLP, speech recognition, and other AI fields [184]. Compared to conventional ML methods, DL techniques can handle larger and more complex datasets and can result in improved performance on certain tasks [185]. DL algorithms build complex models by stacking multiple neural network layers, which are called deep neural networks. Pre-training is a DL model training strategy, in which models are initially trained on a large-scale data set to learn a common feature representation that is suitable for application in a range of different scenarios. The pre-trained models are subsequently fine-tuned to achieve optimal results when applied to specific tasks. As shown in Table 3, DL approaches have been widely used in both sentiment/emotion analysis and misinformation detection. For example, Iwendi et al. [21] explore the use of RNN, GRU, and LSTM as classifiers to detect fake news relating to COVID-19, based on 39 features (including sentiment, linguistic features, and named entities) extracted from news articles and social media posts. Ajao et al. [153] apply various machine learning methods and an LSTM with hierarchical attention networks (HAN) [186] for rumor detection. A Bi-LSTM is used by Hamed et al. [173] to detect misinformation using dual emotions and content features. In [20], the authors adopt CNN and Bi-GRU to extract dual emotion features. To evaluate the effectiveness of their proposed multitask framework for rumor detection, Choudhry et al. [30] employ various DL methods, including LSTM, BERT, CNN, RoBERTa, CapsuleNet [187], and HAN. Various studies apply GCN and Graph neural network (GNN) to model the graph-like structure of social media posts [163, 179].

Pre-trained models, including BERT, DistilBERT, and RoBERTa, are frequently used as the basis for extracting sentiment and emotion features in the context of misinformation detection [22, 24, 161, 163, 169, 176, 178, 180]. A popular technique is to use transfer learning to fine-tune these pre-trained models on large emotion detection datasets (e.g., GoEmotions [188] and DailyDialogue [189]) prior to labeling misinformation datasets. Moreover, there exist a small number of pre-trained models for languages other than

Table 2
Summary of misinformation datasets. A: Available, N: No link, R: Request. An empty cell in the *Notes* column means that the dataset is in English and consists only of textual data.

Dataset	Source	Description	A	Notes
PHEME [104]	Twitter	105,354 tweets organized into 6425 threads (2402 rumors and 4023 non-rumors), relating to nine events. (A thread consists of tweets introducing a news item and a series of follow-up comments)	A	
FakeNewsAMT[111]	various	240 fake and 240 legitimate news items	A	
Celeb [111]	various	250 fake and 250 legitimate news items in the celebrity domain	A	
Twitter15 [105]	Twitter	1490 source tweets (374 non-rumors, 370 false rumors, 372 true rumors, 374 unverified rumors) with retweets and replies	A	
Twitter16 [105]	Twitter	818 source tweets (205 non-rumors, 205 false rumors, 205 true rumors, 203 unverified rumors) with retweets and replies	A	
Twitter16-2[106]	Twitter	498 rumors and 494 non-rumors with comments	A	
ISOT [112, 113]	various	23481 fake and 21417 genuine news items with titles from 2016-2017, focused on political and world news topics	A	
LIAR [114]	Politifact	12.8k manually labeled short claim statements in various contexts with speaker related meta-data, primarily from 2007-2016	A	
LIAR-PLUS [115]	Politifact	An extended version of the above LIAR dataset, in which the claims are accompanied by sentences that provide justifications for the assigned labels	A	
CREDBANK [116]	Twitter	60 million tweets from 2014-2015, concerning various topics grouped into 1049 real-world events, each labeled by 30 human annotators	A	
Kaggle Fake News dataset [117]	various	12,999 posts, consisting of both text and metadata, collected over a period of 30 days from 244 websites	A	
George McIntire dataset	various	6.3k news items, with an equal distribution of fake and real items. (https://github.com/GeorgeMcIntire/fake_real_news_dataset)	A	
SLN [118]	various	360 news articles covering 12 contemporary news topics in 4 domains (civics, science, business, and soft news)	A	
LUN [119]	various	News items classified as trusted(13995), satire(14985), hoax(12047) or propaganda (35029)	A	
Twitter_harvard[120]	Twitter	111 events with tweet ids and user information (60 rumors and 51 non-rumors)	A	
Health_related news [121]	Twitter	709 posts (54% rumour, 30% non-rumour and 16% unknown), collected using the keywords *#zikavirus* and *zika microcephaly*	R	
MultiSourceFake[122]	various	5,994 real and 5,403 fake news articles	A	
PoliticalNews [123]	various	14,240 news pages from 2013- 2018 (7,136 fake and 7,104 genuine)	A	
Buzzfeed Political News [124]	various	Dataset 1 (Buzzfeed 2016 election data): 36 real and 35 fake items; Dataset 2 (political news): 75 real, 75 fake and 75 satire items; Dataset 3 (Burfoot and Baldwins satire): 233 satire and 4000 real items	A	
[125]	various	23,935 news items from September 1995 to January 2021	A	
HWB [126]	various	500 real and 500 fake documents related to health and well being	A	
[127]	various	News articles from eight web sources concerning the Hanoi summit between the presidents of the United States (Donald Trump) and North Korea (Kim Jong-un)	N	
MultiFC [128]	various	36,534 multi-domain claims with their metadata (different domains have different labels, which encompass both direct truth ratings ("correct," "incorrect") and labels that are difficult to map to a level of truthfulness (e.g. "grassroots movement!", "misattributed", "not the whole story"))	A	
RAWFC [129]	Snopes	2,012 claims with supporting evidence, labeled either true, false, and half-true.	A	
FND23 [130]	various	833 news articles that were published after 2020	N	
Infodemic [100]	various	10,700 social media posts and articles (5600 real, 5100 fake) about COVID-19.	A	COVID-19
CoAID [101]	various	4,251 news items (204 fake and 3,565 true news articles, 28 fake and 454 true claims), 296,000 related user engagements (e.g. clicks, shares), 926 social platform posts about COVID-19.	A	COVID-19
[102]	various	586 genuine and 578 fake news items, and more than 1,100 news items and social media posts regarding COVID-19.	A	COVID-19
[103]	Twitter	Globally-collected Tweets related to the COVID-19 epidemic, obtained by filtering tweets containing word or hashtag *Covid-19, Corona Virus, Corona, COVID, covid19*, and *sarscov2*	R	COVID-19
LTCR [131]	various	1,729 real and 500 fake news items related to COVID-19	A	COVID-19, Chinese
FakeNewsNet[108]	Politifact	432 fake and 624 real news items with content, images, and social network information	A	multimodal
FakeNewsNet[108]	Gossipcop	5,323 fake and 16,817 real news items with content, images, and social network information	A	multimodal
Fakeddit [109]	Reddit	1,063,106 samples with submission title, image, comments, and metadata	A	multimodal
MediaEval2016 [110]	Twitter	193 cases of real and 220 cases of misused images/videos, associated with 6,225 real and 9,596 fake posts posted by 5,895 and 9,216 unique users, respectively	A	multimodal
NovEmoFake [132]	various	6816 real and 4950 fake news items (text and images) with background information (where and in which context the news item was first published)	R	multimodal
MMM [133]	various	5630 real and 4840 fake news items (text and images) with background information (where and in which context the news item was first published)	A	multimodal,Hindi, Bengali, Tamil

Dataset	Source	Description	A	Notes
Multimodal-Weibo[134]	Weibo	9528 posts (4749 rumor and 4779 non-rumor) with images, created between May 2012 and January 2016	N	multimodal, Chinese
Weibo21 [107]	Weibo	4,488 fake and 4,640 real news items from 9 different domains collected between December 2014 and March 2021 with news text, image content, timestamps, and comments	A	multimodal, Chinese
Weibo16 [106]	Weibo	2313 rumors and 2351 non-rumors with comments	A	Chinese
Weibo16 (deduplication) [29]	Weibo	3706 news items (1,355 fake, 2351 real) with comments	A	Chinese
Weibo20 [29]	Weibo	6362 news items (3161 fake, 3201 real) with comments	A	Chinese
Weibo20-miao [23]	Weibo	3034 rumors and 3034 non-rumors created between 2016 and 2020	A	Chinese
[135]	Weibo	7880 fake and 7907 real news items with approximately 160k comments	N	Chinese
Portuguese-data[136]	various	76,782 news items, prelabeled according to whether they were sourced from true or fake news sites	N	Portuguese
FakeNewsSet[137]	Twitter	300 fake and 300 genuine news items	A	Portuguese
Fake.Br [138]	various	3,600 fake and 3,600 genuine news items classified into six categories (politics, TV & celebrities, society & daily news, science & technology, economy, and religion)	A	Portuguese
CLEF2020[139]	competition	Five tasks related to verification of claims: task1: check-worthiness of tweets (962 tweets in English and 7,500 tweets in Arabic); task 2 - verified claim retrieval (1,197 tweets and 10,375 verified claims in English); task 3 - evidence retrieval (200 claims and 14,742 corresponding Web pages containing evidence).; task 4 - claim verification (165 claims in Arabic); Task 5 - check-worthiness of debates (70 debate transcripts in English).	A	English, Arabic
[28]	Twitter	202 false rumors and 201 true rumors relating to 403 real-world events with comments	R	Arabic
DAST [140]	Reddit	3007 source posts (273 Support, 300 Deny, 81 Query, 2353 comments); 3007 top-level comments (261 Support, 632 Deny, 304 Query, 1810 comments)	A	stance, Danish
ByteDance fake news dataset [141]	ByteDance	320,767 news pairs in both Chinese and English; test data contains 80,126 news pairs. Given the title of a fake news article A and the title of an incoming news article B, participants are asked to classify B according to whether it agrees with A, disagrees with A,or is unrelated to A	A	stance, Chinese and English
RumourEval17 [33]	Twitter	Task A (stance classification): 5568 posts (1004 Support, 415 Deny, 464 Query, 3685 comments); Task B (veracity prediction): 325 source posts (145 True, 74 False, 106 Unverified) with associated comments	A	stance
RumourEval19 [142]	Twitter, Reddit	Task A (stance classification): 8574 posts (1184 Support, 606 Deny, 608 Query, 6176 comments); Task B (veracity prediction): 446 source posts (185 True, 138 False, 123 Unverified) with associated comments	A	stance
FNC [143, 144]	Snopes, Twitter	49972 tuples, each consisting of a headline-body pair	A	stance
Covid-Stance [145]	Twitter	14,374 tweets (2848 Neutral, 4685 Against, 6841 Favor) related to COVID-19	A	stance
Emergent [143]	various	300 claims and 2,595 associated article headlines	A	stance
PHEME_stance [146]	Twitter	297 threads containing 4,561 tweets (including retweets), spanning 138 rumors organized into 9 stories	A	stance
London Riots [147]	Twitter	7297 tweets concerning 7 different rumors (5761 support, 957 deny, 579 question)	N	stance
[148]	Twitter	327,484 tweets concerning 72 rumors (60.9% support, 27.4% against)	N	stance
2020 US Presidential Election [149]	Twitter	2500 tweets manually labeled with stance, 1250 for each presidential candidate (Joe Biden and Donald Trump)	A	stance
Sydney Siege [150]	Twitter	4375 tweets (2906 affirm, 1469 deny)	N	stance

English, such as AraBERT-Twitter and MARBERT, which have been used for rumor detection in Arabic social media [28].

For multimodal data, ResNet [190], VGG [191], and Xception [192] have been used to extract image features [132, 133, 167, 168]. Similarly to text-only datasets, these studies use transfer learning to fine-tune pre-trained image models on visual sentiment datasets, then extract image features from the misinformation dataset, and finally merge them with text features to perform misinformation detection.

4.4. Advanced Fusion Methods

A wide variety of methodologies for emotion-based misinformation have been developed (See Table 3 for a complete list). In the majority of cases, information about emotions and/or sentiment is fused with other types of features, aiming to take full advantage of the specific characteristics of the dataset used to maximize detection performance. Additional features may be based, for example, on various aspects of textual content; information regarding the structure or temporality of collections of social media posts; and/or images associated with textual data. Moreover, approaches vary in terms of whether they carry out learning within the context of a single or multitask framework. In this section, we introduce a selection of these advanced fusion methods, which are categorized according to the types of additional features and/or the learning strategy that they employ.

Table 3
Summary of emotion-based misinformation detection methods. AF: Affective Features, E: Emotion, S: Sentiment, IE: Image Emotion, ED/SA: Emotion Detection/Sentiment Analysis, ERD: Emotion-based Rumor Detection, MLs: Various Machine Learning methods. Other abbreviations are explained in section 4.2.

Pub	Year	Data	AF	ED/SA	ERD	Other Features
[135]	2019	Customized	E	GRU	EFN (Fig. 6 (c))	
[151]	2019	PolitiFact, GossipCop	S	VADER	SAME	Image, User-based
[152]	2019	Various Public Dataset	S	NB	NB, RF	tf-idf scores, Cosine similarity scores
[153]	2019	PHEME	S	LIWC	MLs, LSTM-HAN	Topics
[154]	2019	Weibo16	E	ALO	SD-DTS-GRU (Fig. 9 (b))	Time
[155]	2020	Weibo16, Twitter16	S, E	Dictionaries	SD-TsDTS-CGRU (Fig. 9 (b))	Time
[156]	2020	PHEME, Twitter15, Twitter16	S		a Hierarchical Attention Network with User and Sentiment information	User-based
[126]	2020	HWB	E	NRC Intensity Emotion Lexicon	MLs	
[157]	2020	Fake.Br	S, E	Dictionaries	MLs	Grammatical, Stylistic
[158]	2020	FakeNewsNet, CredBank	S	Dictionaries	DT, Bi-LSTM	Topics
[159]	2020	CLEF2020	S	Dictionaries	Web Check	Topics, Offense Named Entities
[160]	2020	FakeNewsNet	S	VADER	MLs, DNN	Retweet Rate
[22]	2021	Weibo16	S	BERT	TDEI (Fig. 9 (a))	Time, Propagation Structure
[161]	2021	ByteDance, FNC, Covid-Stance	S, E	BERT, LSTM	Multitask (Fig. 10 (b))	Textual Novelty
[29]	2021	RumourEval19, Weibo16, Weibo20	S, E	Dictionaries	MDE (Fig. 6 (b))	Dual Emotion
[162]	2021	FakeNewsSet	S, E	Dictionaries	MLs	Image Captioning, Grammatical, Stylistic
[23]	2021	Weibo16, Weibo20	S	ALO, BERT	SSE-BERT (Fig. 7 (a))	Dependency Tree
[122]	2021	MultiSourceFake, LUN, PoliticalNews, FakeNewsNet	S, E	Bi-GRU basd on dictionary	FakeFlow (Fig. 5 (a))	Topics
[30]	2022	PHEME, FakeNews AMT, Celeb, Gossipcop	E	Unison model	Multitask (Fig. 10 (a))	Domains
[24]	2022	ISOT	E	RoBERTa	RoBERTa, RF	
[21]	2022	[102]	S		GRU, LSTM, RNN	Stylistic, Linguistic-informed
[163]	2022	Twitter15, Twitter16	S	RoBERTa	SA-HyperGAT (Fig. 8 (b))	Structure
[103]	2022	Customized	S	VADER	Modified VADER	Diffused Information
[164]	2022	MultiFC	S, E	EmoAttention BERT	EmoAttention BERT (Fig. 5 (c))	
[165]	2022	Infodemic, CoAID	S	Fuzzy Sentiment Scoring	LSTM (with Fuzzy Sentiment) (Fig. 5 (d))	
[127]	2022	Customized	S	SenticNet	Conceptual Graphs with sentiment	Entity
[133]	2022	MMM [132]	IE	ResNet	SCL, BERT and ResNet	Novelty, Image,
[166]	2022	PHEME	S, E	EmoLex	GCS (Fig. 8 (c))	N-gram, Similarity Matching
[167]	2022	MediaEval2016, Multimodal-weibo	IE	VGG	CredNN	
[168]	2022	Fakeddit	IE	Xception	BERT, Xception	Image, Image caption
[169]	2022	ByteDance, Covid-Stance, FNC, LIAR-PLUS	E	BERT	LR	Novelty
[170]	2022	Weibo16, RumourEval	S	Dictionaries	RvNN with Temporal (Fig. 9 (c))	Time
[171]	2022	Buzzfeed Political News	S	SEO Scout's analysis	MLs	Stylistic, Linguistic-informed
[125]	2022	Customized	S	AFINN, VADER	MLs	Topics, Title-text similarity
[28]	2023	Customized	S, E	Dictionaries	Arabic PLMs	Topics, Stylistic, User-based
[20]	2023	RumourEval19, PHEME, Fakeddit	S, E	CNN, Bi-GRU	AGWu-RF (Fig. 6 (a))	Stylistic
[172]	2023	ISOT, LIAR	S	Lexicon-based	MLs	Grammatical, Likes
[173]	2023	Fakeddit	S, E	Dictionaries	Bi-LSTM	Title
[174]	2023	PolitiFact, GossipCop	S, E	method in [29]	BERT, ResNet	Title-Text similarity,Images
[175]	2023	Weibo16, Twitter15, Twitter16	S	Sentiment Pattern Module (SPM)	ptVAE (Fig. 7 (b))	User-based, Structure, Propagation
[176]	2023	SLN, LUN	S, E	RoBERTa, Sentiment Interaction Graph	MHN (Fig. 8 (a))	Graph, Topics, Entities
[177]	2023	Twitter_harvard, Health_related	E	EmoLex	MLs	Linguistic-informed, User-based

Emotion detection for misinformation

Pub	Year	Data	AF	ED/SA	ERD	Other Features
[136]	2023	Portuguese datsaset (customized), Fake.br	S	Sentiment Gradient	MLs, LSTM	
[132]	2023	NovEmoFake (customized)	IE	ResNet	SCL, BERT and ResNet, VisualBERT	Novelty, Image
[178]	2023	Weibo21	S	BERT	Mixture-of-Experts (Fig. 5 (b))	
[179]	2023	Weibo16	S	Bi-GRU with Attention model	Bi-GCN	Semantic, Propagation information
[180]	2023	[181]	S, E	RoBERTa, DistilBERT	An Ensemble Method, RNN	
[182, 183]	2022, 2023	FakeNewsAMT, Celeb, Politifact and Gossipcop datasets	E	Unison model	Multitask (Fig. 10 (a))	Domains

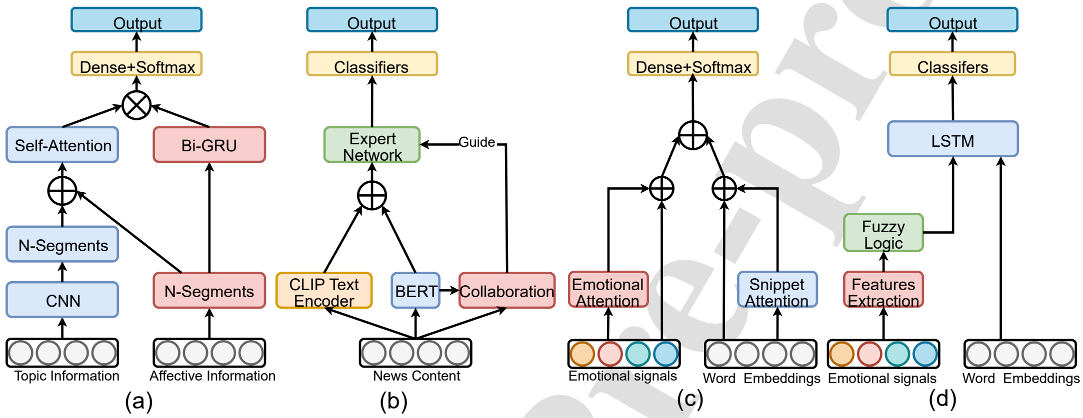


Figure 5: Emotion-based misinformation detection by combining emotion with other text-based features. (a) FakeFlow [122], (b) Mixture-of-Experts[178], (c) EmoAttentionBERT[164] (d) LSTM (with Fuzzy Sentiment)[165]

4.4.1. *Methods Combining Emotion with Other Text-Based Features*

Various methods have attempted to exploit the wealth of valuable information conveyed in text by combining emotion/sentiment features with other features derived from the textual content of news articles or social media posts.

Ghanem et al. [122] propose the FakeFlow model (Figure 5 (a)), which aims to model the flow of affective information in fake news articles, based on the hypothesis that the pattern of affective information in fake news differs from that found in genuine news, e.g., emotions of fear are often evoked towards the start of fake new articles. The model consists of two modules, the first encoding topic information, extracted using a CNN, and the second capturing 23 affective features relating to emotion, sentiment, morality, imageability, and hyperbola. In the first module, potential relationships between topics and affective information are captured by concatenating their respective vectors (e.g., emotions in a fake article about Islam are likely to vary from those in an article in favor of a politician). A context-aware self-attention mechanism is subsequently applied to weight segments according to their similarity to neighboring segments. In the second module, the flow of the affective information within the articles is modeled by feeding the affective vectors to Bi-GRU. Finally, a dot product and average operation are applied to distill the output of the two modules into a compact representation, which is fed into a fully connected layer and a softmax layer to determine the factuality of the article.

The multi-domain fake-news detection system described in [178] (Figure 5 (b)) is based on mixture-of-experts model, which involves training multiple neural networks based on TextCNNs (experts), each targeted at a different part (domain) of a dataset. Pre-trained BERT and CLIP [193] text encoders are applied to obtain two different embeddings of news content, which are combined as a fusion embedding. The use of the CLIP text encoder, which is pre-trained on image-text paired datasets, aims to take advantage of the rich semantic representations obtained through state-of-the-art (SOTA) multimodal learning. A *Collaboration* module adaptively determines the weights of each expert model, to enhance or suppress their contribution in the final mixture-of-experts model. The module consists of a fusion vector, which combines sentence-level embeddings e_a from attention, sentiment embeddings e_s obtained by fine-tuning BERT using the Weibo_senti_100k dataset[5], and domain embeddings e_d. The expert networks are accumulated and multiplied via the collaborative influence function C_i, which is determined by the Collaboration module, and then used for classification.

[5] https://ieee-dataport.org/documents/weibosenti100k-and-thucnews

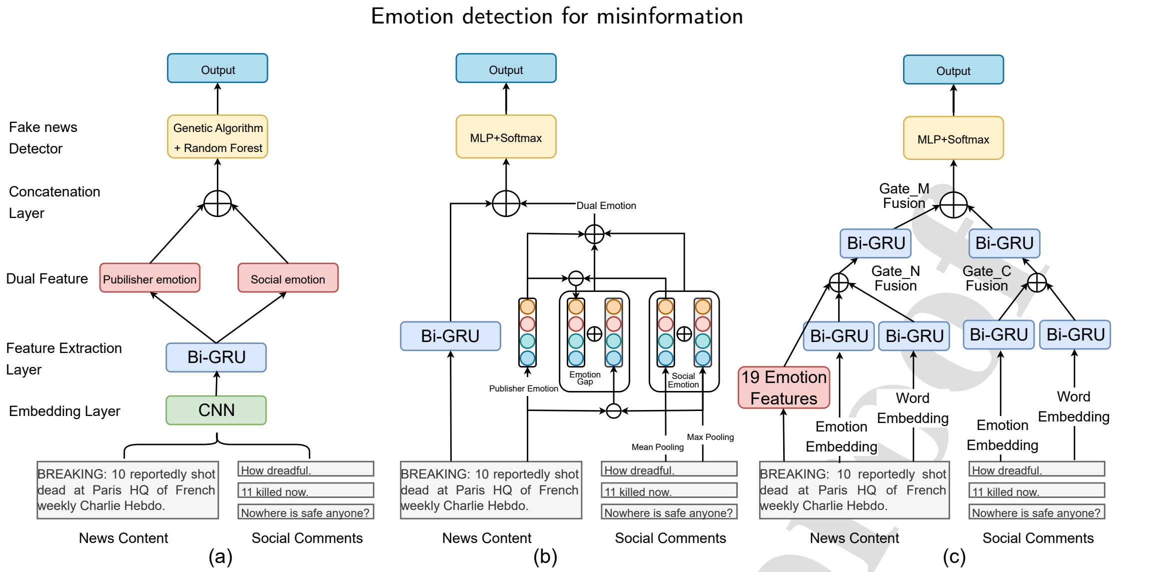


Figure 6: Emotion-based misinformation detection by mining of dual emotions. (a) AGWu-RF[20], (b) MDE[29], (c) EFN[135]

The EmoAttention BERT model architecture [164] (Figure 5 (c)) uses both emotion and snippet attention to verify the truth of political claims, supported by evidence from Google news snippets. The content of snippets is encoded using word embeddings, while the NRC Intensity Emotion Lexicon [194] is used to calculate word–level intensities for eight basic emotions. An emotional attention layer assigns a weight to each emotion vector to identify the most relevant emotional signals in a given evidence snippet, while a snippet attention layer weights each evidence snippet with respect to the associated claim. Finally, the vectors from both layers are distilled and fed into a softmax layer to predict the truth of the claim.

Mohamed et al. [165] detect fake news using an LSTM that combines textual embeddings from Word2Vec [195] with fuzzy sentiment features (Figure 5 (d)). Sentiment features are extracted by firstly identifying opinion-denoting words and associated polarity information using SentiWordNet [196] and WordNet[6], resulting in an initial score. These values are subsequently modified using fuzzy logic functions, according to the presence of different types of linguistic hedges (i.e., words that modify the intensity and meaning of an expressed opinion, such as *not, very,* and *quite*), using fuzzy logic functions to obtain the final sentiment score.

4.4.2. Mining of Dual Emotions

A number of studies have investigated how misinformation detection in social media can be improved by taking into account information about the different emotions expressed in posts that announce news (i.e., *publisher posts*) and posts that comment on or react to these source posts (i.e., *social posts*).

Luvembe et al. [20] propose a deep normalized attention-based mechanism for enriched extraction of dual emotion features (Figure 6 (a)), which combines CNN and Bi-GRU. The CNN layer is used to obtain embeddings for both publisher and social posts, after which a stacked Bi-GRU with attention is utilized to extract and concatenate emotion features from each type of post. Classification of publisher tweets according to whether or not they report misinformation is performed using an RF model, whose features are guided by a genetic algorithm, which determines the subset of features that can achieve optimal classification performance.

The MDE model [29] (Figure 6 (b)) detects misinformation in social media posts by integrating features from existing Bi-GRU fake news detectors with publisher and social emotion features, and the relationship between them. A vector representing emotions in the publisher post emo_P, is obtained by concatenating the emotion category, lexicon-based emotion score, emotional intensity, sentiment score, and other auxiliary features (e.g., emoticons and punctuation). A vector is created for each social post, by applying the same method used to obtain the publisher emotion vector. The individual social emotion vectors are subsequently combined, after which they are aggregated in two ways, i.e., using mean pooling emo_S^{mean} (to represent average emotion signals) and max pooling emo_S^{max} (to capture extreme emotional signals). These two types of aggregation are then concatenated to obtain the overall *Social Emotion* emo_S. The *Emotion Gap* emo_{gap} represents the difference between the publisher and social emotions, and is obtained by concatenating (emo_P - emo_S^{mean}) and (emo_P - emo_S^{max}). Finally, dual emotion features are obtained by concatenating the publisher emotion (emo_P), the social emotion (emo_S), and the Emotion Gap (emo_{gap}). These features are combined with those from the existing Bi-GRU fake news detector, and fed into a multi-layer perceptron (MLP) layer and a softmax layer to determine whether or not the publisher post represents fake news.

[6]https://wordnet.princeton.edu/

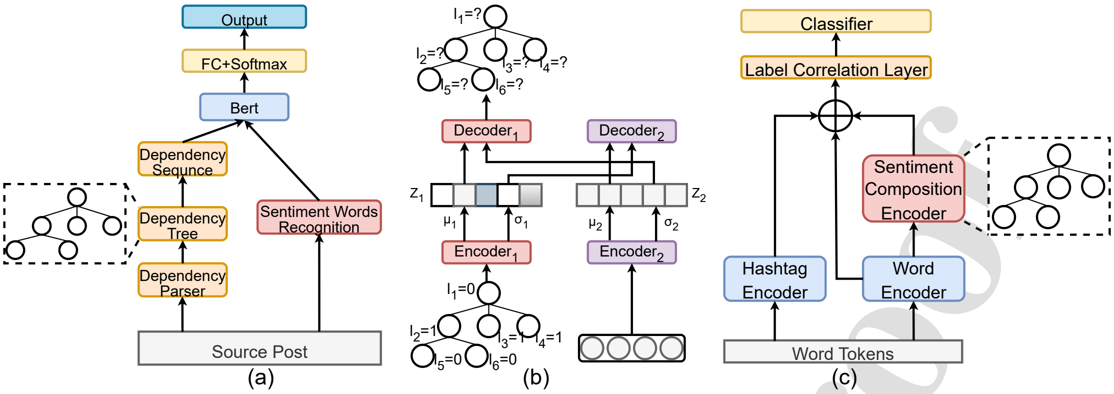


Figure 7: Emotion-based misinformation detection based on tree structures. (a) SSE-BERT[23], (b) ptVAE [175], (c) Multi-EmoBERT[86]

The end-to-end emotion-based fake news detection framework for social media (EFN) proposed by Guo et al. [135] (Figure 6 (c)) consists of a content module, comment module and fake news detection module. The content module (left of the figure) is used to encode publisher posts using Bi-GRU for word embeddings and emotion embeddings, the latter of which is trained using large-scale Weibo datasets, with emoticons as the emotion labels. A gate recurrent unit (Gate_N) is then applied to combine word embeddings, emotion embeddings, and 19 sentence-based emotion features. Subsequently, all vectors are fed into another Bi-GRU, whose final hidden state is used as the representation of the publisher post. The comment module (right of the figure) represents information about follow-up social posts. The comment module architecture is similar to that of the content module, except that all comments are concatenated before being fed into the Bi-GRU, and sentence-based emotion features are not used. A different gate (Gate_C) is used to fuse features. Finally, the output of the third gate (Gate_M), which combines the content and comment representations, is fed to a fully connected layer with softmax activation to determine whether or not the publisher post constitutes fake news.

4.4.3. Methods Based on Tree or Graph Structures

Due to the inherent relationships among posts relating to fake news, such as retweets or likes of source posts from followers on Twitter, social media data may be viewed as tree or graph structures, through which information propagates. Accordingly, several methods employ such structures to model the spread of information and to capture relationships between nodes of the tree. The words and phrases that make up sentences can also be arranged into hierarchical tree-like structures, according to the grammatical and semantic relationships that hold between them. Several misinformation detection methods make use of features based on these relationships, including dependency tree and sentiment tree information.

In [23], an *earliest rumor detection* approach for social media is described (Figure 7 (a)). It considers only publisher posts without their follow-up social comments, with the aim of catching potentially harmful rumors before they become widespread. The use of a syntax and sentiment-enhanced version BERT (SSE-BERT) is inspired by the observation that both the sentiment *and* syntactic features of rumors are often distinct from non-rumors. Syntactic dependency trees are firstly obtained for each source post using DDParser [197], and are encoded into a dependency sequence by preorder traversal. Sentiment-denoting words belonging to seven different categories are subsequently recognized using an external sentiment lexicon (i.e., ALO [198]). Specific embeddings are then assigned to each token according to the sentiment lexicon. All features are learned by BERT and distilled using element-wise addition. Finally, the vector of [CLS] in BERT is fed into a fully connected network with softmax to detect rumorous publisher posts.

Driven by the scarcity of high-quality annotated training data, [175] proposes an unsupervised approach, ptVAE (Figure 7 (b)). Based on the observations that rumorous tweets exhibit different sentiment patterns compared to rumorous tweets, and that they diffuse more rapidly, deeply and broadly, the method aims to detect rumors by identifying collections of tweets whose propagation patterns and sentiment characteristics differ from those of normal (i.e, non-rumorous) collections. The proposed model consists of a Sentiment Pattern Module (SPM), Propagation Feature Module (PFM), and Cross-alignment module. In the SPM (left of Figure 7 (b)), a tree encoder infers the pattern of sentiment labels along the input propagation tree and uses a GRU to encode this pattern into a latent vector z_1. The original sentiment labels for each node are then reconstructed by decoding z_1 using a node label decoder and a child label distribution decoder which, respectively, predict the label of each node and the label distribution of the node's children. The PFM (right of Figure 7 (b)) creates vectors capturing the speed, and depth & breadth of propagation, and combines them as the input to a VAE, whose encoder and decoder

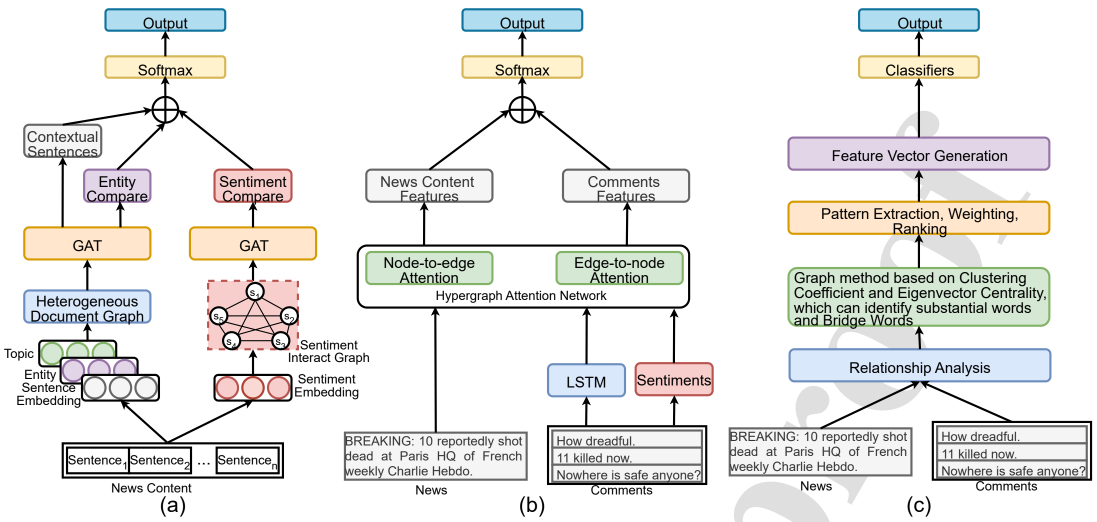


Figure 8: Emotion-based misinformation detection based on graph structure. (a) MHN[176], (b) SA-HyperGAT [163], (c) GCS[166]

are based on a multilayer perception. The Cross-alignment module then jointly learns the propagation tree of the SPM and the propagation characteristics of the PFM.

Li et al. [86] propose the Multi-EmoBERT model (Figure 7 (c)) to detect multiple co-existing emotions in fake news content on social media platforms. The first part consists of a word encoder to obtain representations of words, and a hashtag encoder to obtain representations of emotion-word hashtags and emojis, which are common features of social media text. The second part is the sentiment semantic composition encoder, which uses the Stanford CoreNLP toolkit to construct sentiment trees, and employs a self-attention mechanism and phrase node selection to obtain phrase level vectors. The final part is a label correlation layer, which uses a parameter to capture correlations between co-existing emotions. A subsequent analysis, revealing that multiple emotions are often conveyed within a single fake news posting, demonstrates the potential value of Multi-EmoBERT in detecting fake news posts.

The method described in [179] combines the use of semantic and sentiment information, along with the structure of information propagation in social network posts, to obtain enriched features for rumor detection. BERT is used to separately capture information about publisher tweets and follow-up social comments, while a Bi-GRU with attention is used to encode sentiment information conveyed in follow-up tweets. Propagation features of tweets are obtained with the aid of a Bi-GCN network. The various features are then spliced and fused to detect rumors.

Figure 8 (a) illustrates the graph attention Mixed Heterogeneous Network (MHN) based model developed by Zhang et al. [176], aimed at detecting longer news articles that contain misinformation. The approach is based on the finding that patterns of sentiments expressed across sentences in fake news articles are usually very different from patterns in real news articles. The model employs two types of graph structures. Firstly, a sentiment interaction network encodes sentence-level sentiment features using a pre-trained RoBERTa model, and captures changes in sentiment that occur in the context of the surrounding sentences. A sentiment comparison model calculates comparison vectors between each contextual sentiment representation obtained from the input news document and its corresponding original sentiment embedding; discrepancies between these embeddings could be indicative of fake news. Secondly, a heterogeneous document graph encodes the semantic content of the article, by capturing interactions between sentences, topics, and entities. A comparison between the contextual entity vectors and those obtained from a knowledge graph is aimed at detecting potential information inconsistencies that could denote fake news. The sentiment comparison vector, entity comparison vector, and article representations are combined and passed through a softmax layer to make predictions.

Dong et al. [163] design the Sentiment-Aware Hypergraph Attention Network (SA-HyperGAT) for fake news detection in social media (Figure 8 (b)). Compared to general graphs, the use of hypergraphs is intended to capture higher-order dependency information between words and sentences. Separate hypergraphs are constructed for publisher posts and follow-up social comments. In the former hypergraph, each node corresponds to a word in the news text, while in the latter, each node corresponds to a user comment. Sentiment labels for each comment, obtained using a fine-tuned RoBERTa model, are used as hyperedges in the graph. Representations of comments are learned using an LSTM, after which node-to-edge attention and edge-to-node attention are applied to learn the representations of the hypergraphs. Final feature vectors are obtained by applying mean pooling to both hypergraphs; these vectors

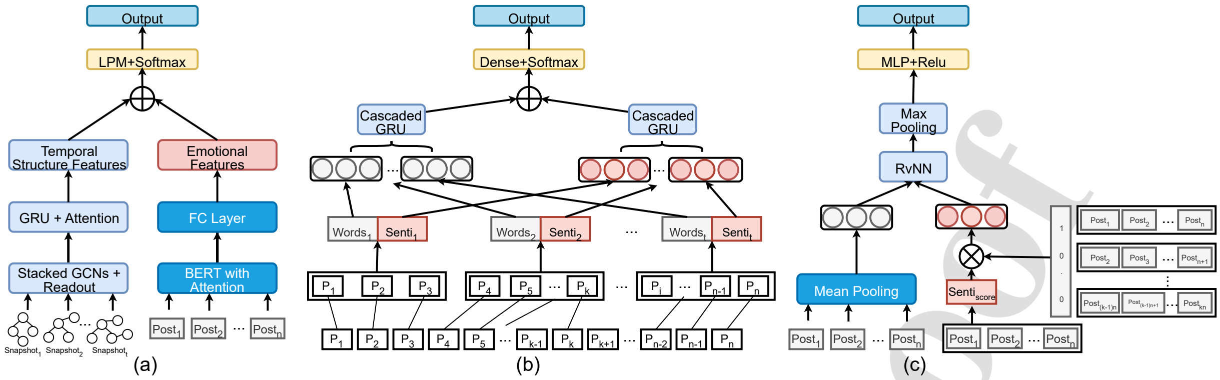


Figure 9: Emotion-based misinformation detection based on temporal information. (a) TDEI[22], (b) SD-TsDTS-CGRU[154, 155], (c) RvNN with Temporal [170]

are combined and then fed into a softmax classifier to obtain the final prediction.

The graph-based contextual and semantic learning (GCS) method for detecting rumors in tweets [166] (Figure 8 (c)) is based on a novel approach to graph-based representation learning, and the identification of two prevalent categories of words that constitute the building blocks for constructing contextual patterns for rumor detection. These categories are *substantial words*, which are used to express emotions, sentiments, or suspicions about the event, and *bridge words*, which connect substantial words. After data pre-processing, publisher and social tweets are combined to allow important relationships between them to be identified, e.g., social tweets may convey skepticism, correction, verification, etc., towards the publisher tweet. The combined tweets are represented as word co-occurrence graphs, to which *clustering coefficient* and *eigenvector centrality* are applied to identify substantial and topical words, and bridge words, respectively. These are further enriched with negative emotion patterns and skeptical patterns. Next, a modified TF-IDF formula is used to rank and select the top-k patterns that are most likely to be indicative of rumor. Semantic vectors are then generated for both tweets and patterns using word embeddings, which are combined and converted into features using cosine similarity, for subsequent use by different conventional machine classification algorithms (i.e., SVM, GB, conditional random field, and LR).

4.4.4. Methods Based on Temporal Information

Various temporal features have been explored to enhance the performance of misinformation detection, based on the time-sensitive patterns that are frequently observed in social media, e.g., rumors initially spread quickly but gradually disappear, while reader emotions tend to change over time.

The TDEI model [22] (Figure 9 (a)) integrates emotion features with information concerning time-sensitive dynamic changes in the topological propagation structure of tweets, which is considered to be a better predictor of rumors than the final, static propagation structure. The graph representing the propagation structure of a publisher post and its associated social comments is firstly divided into a sequence of temporal snapshot graphs. Stacked GCNs and a readout function are used to learn the structural features of the snapshots. A GRU with self-attention is then applied to learn the diffusion process of structures. Meanwhile, emotion vectors are extracted from each post using a pre-trained, fine-tuned BERT model. A self-attention mechanism is then used to merge the emotion vectors for each post corresponding to a rumor event into a single vector, whose dimensionality is adjusted using a fully connected layer. The temporal dynamic structure and emotion vector are then concatenated and fed into multi-layer perception with softmax function to make predictions.

The SD-TsDTS-CGRU fusion rumor detection method [154, 155] (Figure 9 (b)) focuses on detecting rumors at the event level, i.e., by considering all information expressed in the complete set of sequential posts relating to the same topic or event. Posts are firstly automatically partitioned into sets covering distinct events by dividing them into intervals using a two-step dynamic time series division algorithm, based on fuzzy clustering and information granules [199]. The latter step helps to ensure that each batch of posts covers information at an appropriate level of granularity and has a consistent semantic interpretation. The calculation of information granularity takes into account the number of sentiment words belonging to each fine-grained sentiment category in each interval, obtained using a novel sentiment dictionary containing sentiment words and emoticons. Following the division, word embeddings and sentiment information extracted from the posts in each event-specific set are fed into two different GRUs, whose outputs are combined and fed into a dense layer with softmax function to predict whether or not each set of event-related posts constitutes a rumor.

Temporal sentiment features of rumors are employed in [170] (Figure 9 (c)) to account for changes in sentiment over the lifetime of an original publisher post and its associated social posts, in both Chinese and English social media

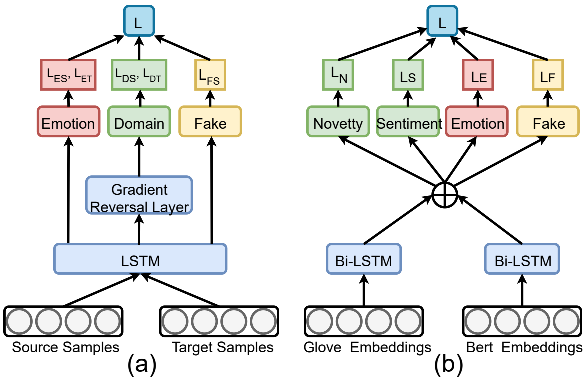


Figure 10: Emotion-based misinformation detection based on multitask learning. (a) [30, 182, 183], (b) [161]

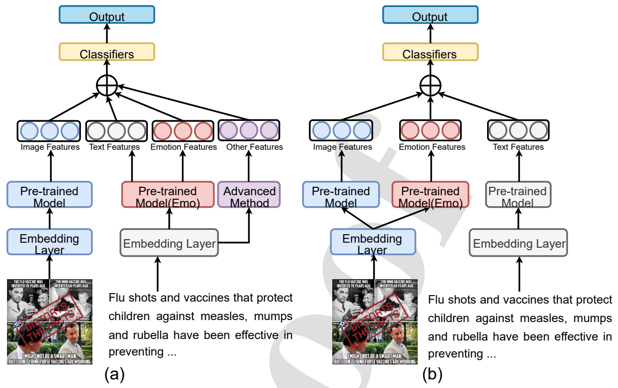


Figure 11: Emotion-based misinformation detection based on temporal information. (a) Multimodel with text emotion. (b) Multimodel with image emotion

datasets. The Baidu sentiment API[7] and NLTK sentiment module[8] are used to obtain sentiment scores for Chinese and English posts, respectively. Temporal features are characterized using a one-hot vector, whose length is modified by normalizing the number of posts in the reply series. Temporal sentiment features are obtained by multiplying the modified one-hot vector by the sentiment score. Textual features of posts are obtained using pre-trained word embeddings and a mean pooling layer, which are combined with the sentiment vector to derive a microblog representation. An RvNN and max pooling layer are then used to obtain a comprehensive representation of an event as it propagates through the path of social replies, which is passed to an MLP with ReLU to determine whether or not the publisher post is a rumor.

4.4.5. *Multitask Learning*

Multitask learning optimizes several learning tasks simultaneously, exploiting shared information to improve the prediction performance of the model for each task. Auxiliary tasks can be added to the main task to boost the performance. Several studies have explored how emotion and sentiment detection can act as auxiliary tasks in a multitask learning framework to enhance misinformation detection accuracy.

The method developed by Choudhry et al. [30, 182, 183] (Figure 10 (a)) aims to address the issue of cross-domain robustness in determining the veracity of news articles. Generalizability of the method across different domains is achieved using a domain-adaptive framework, whose aim is to facilitate the extraction of domain-invariant features by aligning the source and target domains in the feature space. The multitask learning setup trains an emotion classifier as an auxiliary task in parallel to a fake news detector, to try to improve the alignment between the source and target domains, while adversarial training helps to make the model robust to outliers. The emotion classifier assigns emotion labels according to Ekman's or Plutchik's emotions, with the aid of the Unison model [200]. An LSTM is used as the feature extractor, which is trained using the accumulated loss from the fake news detector, emotion classifier and a domain classifier, the latter of which acts as a discriminator in learning domain-invariant features.

Based on the relatedness between the tasks of detecting fake news, novelty, emotion, and sentiment, Kumari et al. [161] (Figure 10 (b)) propose a multitask learning framework, in which the latter three of these are treated as auxiliary tasks. Using premise-hypothesis pairs as input, the model detects whether or not the hypothesis is fake with respect to the premise. Pre-trained and/or fine-tuned models are firstly used to determine whether the hypothesis is novel with respect to the premise, and whether or not the hypothesis and premise differ in terms of binary emotion values (i,e., sadness/joy/trust vs. anger/fear/disgust/surprise) and sentiment (positive or negative). Different Bi-LSTMs that use pre-trained GloVe and BERT-based embeddings are employed to obtain two different input textual representations, which are concatenated and used as the input to the three auxiliary tasks and the main task of fake news detection.

4.4.6. *Multimodal Methods*

On platforms like Twitter or Weibo, people often attach images to their textual posts to better express their opinions or emotions. Several studies have thus attempted to exploit information from these images to improve the detection of rumors. Approaches are mainly based on two different frameworks, both of which involve combining features from text and images, but which differ in terms of whether emotion features are extracted from text (Figure 11 (a)) [151, 174] or images (Figure 11 (b))[132, 133, 168].

The Title-Text Similarity and Emotion-Aware Fake News Detection method [174] applies BERT with a fully connected layer and ResNet [190] to obtain textual and visual features, respectively. The publisher emotion extractor from [29] (described above in Section 4.4.2) is used to obtain a range of emotion-based feature values from textual news

[7] https://ai.baidu.com/tech/nlp_apply/sentiment_classify
[8] https://www.nltk.org/api/nltk.sentiment.html?highlight=sentiment#module-nltk.sentiment

content. The scaled dot-product attention mechanism is also used to capture similarities between the title and textual features, based on the observation that authors of fake news may attempt to catch the reader's attention by using titles that are not relevant to the news content. All features are subsequently combined and fed into a fully connected layer with softmax to make predictions.

The SAME multimodal embedding model [151] incorporates user sentiment for fake news detection. Firstly, VGG [191] and CNN are used to represent images, while the text is represented using GloVe and MLP, and profiles (i.e., source, publisher, and keywords) are represented using one-hot encoding. An adversarial learning mechanism is then applied to find semantic correlations between different modalities. A novel hybrid similarity loss method based on the Graph Affinity Metric and Local Similarity Metric is used to incorporate the user's sentiments (i.e., positive, negative, or neutral), which are obtained using the VADER lexicon. Finally, a fully connected layer with softmax is applied as a classifier.

The multimodal framework described in [132, 133] makes use of text and images from source-target pairs, in which the target corresponds to a news item from a fake news dataset, while the source corresponds to background information associated with the target news item, extracted from credible websites. BERT and ResNet [190] are firstly used to encode the text and images of source-target pairs, respectively. The textual and visual features are concatenated to obtain multimodal feature representations. These are encoded using VisualBERT [201], which is designed to capture the rich semantics found in images and their associated text. A novelty detection module then uses these multimodal representations to determine the credibility of the new news (target) with respect to prior verified news (source), using supervised contrastive learning (SCL), such that target representations attract source representations that provide support, and repel them otherwise. The second module pre-trains a neural network to predict image emotion labels using two classes, i.e. joy/love/sadness vs. fear/surprise/anger. All features are then fused and passed to an MLP with softmax to make predictions.

Uppada et al. [168] propose a framework for fake news detection that combines visual and textual features. The architecture, which consists of two fine-tuned Xception [192] models, makes use of the Error Level Analysis (ELA) technique to help to identify digitally altered images. One fine-tuned Xception model is trained on an ELA image dataset to detect editing traces in digital images, while the other is trained on a visual sentiment analysis dataset to determine whether images convey positive or negative sentiments. BERT is then applied to learn contextual knowledge from image captions. The output of the three branches is combined and passed to the fake image classifier.

4.5. Emotion-based Stance Detection in Misinformation

In addition to emotions and sentiment, the stance of readers is also an important factor in affecting rumor diffusion. If somebody supports a piece of fake news, he/she is more likely to reshare it. Emotions can impact upon a person's thinking, judgment, and decision-making, which in turn can influence their stance toward a particular topic. This section introduces methods that use emotion as a feature for stance detection in misinformation.

Most work in this area has been driven by shared tasks, in which a number of teams compete with each other to produce the best results for a given task and dataset. Examples of relevant tasks/datasets include RumourEval17 [33], RumourEval19 [142], and the FNC dataset [143, 144]. Lillie et al. [140] construct a Danish stance-annotated dataset (DAST), consisting of Reddit posts. A number of other publicly available stance-annotated datasets have also been used in various studies [45, 150, 211]. Further details about these datasets are provided in Table 2. Most stance detection methods use conventional ML approaches and combine affective features (detected using simple dictionaries or tools) with a variety of other features. Further information about a range of stance detection methods is shown in Table 4. Some of these methods are extended to perform Rumor Detection based on Emotion and Stance (RDES), in which rumors are detected based on a combination of affective features, stance information and other features.

4.6. Discussion

The analysis above revealed the diverse range of emotion-based methods that have been applied to the problem of misinformation detection, with a focus on the wide variety of designs of advanced fusion methods. In this section, we conduct a comparative analysis to identify the most effective strategies. The top part of Table 5 provides performance statistics for a range of the advanced fusion methods discussed in Section 4.4, accompanied by baseline statistics reported in the corresponding publications, obtained by applying conventional ML or DL methods using the same sets of features. The bottom part of the table reports on the performance of a number of categories of simpler approaches, i.e., conventional ML, DL, RDES and LLM-based methods. Where possible, we also provide the results of ablation experiments, in which sentiment/emotion features are excluded, as a means to assess their impact on overall performance. We firstly compare the overall performance of advanced fusion methods with those of the four other categories of methods listed above. We subsequently compare the results of different advanced fusion methods that have been evaluated using common datasets. Finally, we discuss the impact of using different fusion strategies (i.e., feature fusion and model fusion). Although we discuss possible reasons for varying performance levels, it is important to note that these may influenced by numerous factors, including differences in data processing methods, selection

Table 4
Emotion-based stance detection in rumor and fake news. AF: Affective Features, S: Sentiment, E: Emotion, ED/SA: Emotion Detection/Sentiment Analysis, ESD: Emotion-based Stance Detection, RDES: Rumor Detection based on Emotion and Stance.

Pub	Year	Data	AF	ED/SA	ESD	RDES	Other Features
[150]	2016	Sydney Siege data	S	MetaMind sentimentclassifier API	LR, NB, RF		Stylistic, Twitter metadata, Linguistic-informed
[43]	2017	RumourEval17	S, E	SSWE, ZSWE	Ensemble Classifier	Ensemble Classifier	Linguistic-informed, Stylistic, Tweet metadata, User-based, Semantic, Cluster features,
[202]	2017	RumourEval17	S	VADER	XGBoost		Stylistic, Similarity, Twitter metadata, Grammatical features
[203]	2017	RumourEval17, PHEME	S	Stanford sentiment Tree	DT, RF, KNN		Linguistic-informed, Twitter metadata, User-based, Similarity, Lexical features
[204]	2018	RumourEval17	S	NLTK	Linear SVC, LR, RF, DT, SVM	Linear SVC, LR, RF, DT, SVM	Stylistic, Lexical, Conversation-based, User-based features
[205]	2018	RumourEval17	S		Ensemble Classifier	Ensemble Classifier	Stylistic, Twitter metadata
[40]	2018	FNC dataset	S	EmoLex, NRC-Canada	stacked LSTMs		Linguistic-informed, Topics, Similarity features
[206]	2018	Emergent dataset	S	Stanford sentiment Tree	LR, RF		Grammatical, Stylistic, Structural features
[207]	2018	FNC dataset	S	lexicon based	CNN, LSTM, GRU		Stylistic, Linguistic-informed features
[41]	2019	RumourEval19	S, E	Various Dictionaries	LR	LR	Lexical, Syntactic, Stylistic, Twitter, Conversation-based, Cluster features
[208]	2019	RumourEval19	S	NLTK, VADER	Ensemble Classifier		Stylistic, Linguistic-informed, Grammatical, Semantic, Similarity, User-based features
[209]	2019	RumourEval19	S		Bi-LSTM and rules	Bi-LSTM and rules	Stylistic, Conversation-based, User-based features
[140]	2019	DAST	S	AFINN	LSTM, LR, SVM	HMMs	Stylistic, Lexical, Reddit metadata, Linguistic-informed, Semantic, Similarity features
[44]	2019	RumourEval17	S	SenticNet	LR, DT, RF, LinearSVC, NB		Stylistic, Topic, User-based features
[210]	2019	RumourEval17	E	Various Dictionaries	NB, DT, SVM, RF		Stylistic, Conversation-based
[211]	2020	[148]; London Riots Dataset; PHEME	S	VADER	Graph-based Algorithm		Cluster, Linguistic-informed, Lexical features
[42]	2021	RumourEval19	S, E	Various Dictionaries	Multitask Learning based on longformer	Sentence Encoder	Conversion-based, Stylistic, Grammatical features
[45]	2023	2020 US Presidential Election [149]	S	Topic-based Bi-LSTM [212]	Fuzzy Logic		Semantic, User-based features

of base models, choice of particular SA/ED methods, etc. These factors are discussed further in Section 6 below.

Comparison between advanced fusion methods and other methods: Figure 12 compares the distribution of reported performance scores for advanced fusion methods with the distributions of scores for the four other categories. The results of the baseline methods shown for the advanced fusion methods in Table 5 are included within the distributions of the ML and DL categories. The performance scores in the figure correspond to the best result reported for each detection method in the corresponding publication. It may be observed from Figure 12 that the majority of advanced fusion methods achieve results above 0.9 (40% are above 0.95). Although approximately 30% of the ML methods achieve a score of 0.9 or higher, Table 5 shows that the highest performance scores are generally achieved for customized or highly specialized (e.g., Portuguese) datasets. Overall, DL methods achieve higher results than conventional ML approaches, with the majority of scores ranging from 0.8 to 0.9. In contrast, the scores of most RDES methods fall below 0.6. This could be due to the fact that RDES methods primarily focus on stance detection rather than misinformation detection. Although the performance of LLMs is highly variable, investigations into their use in misinformation detection are still very new. Most of these studies only apply simple designed prompts to test the effectiveness of LLMs or use LLMs as an auxiliary tool for small language models. This helps to explain their unpredictable levels of performance, which can differ significantly according to the evaluation dataset used. As shown in Table 3, the features used in ML and DL methods are similar to those used in advanced fusion methods, and they are also quite diverse. The baseline comparison methods for the advanced fusion methods shown in Table 5 also consist of both ML and DL methods. Apart from the differences in methods, the features and data processing strategies used by both the baselines and

Table 5

Performance of emotion-based methods and LLMs for misinformation detection. *Ablation: No aff. feats*, use the same methods as the *Evaluation* column, but without sentiment/emotion features. The *Evaluation* column provides Evaluation scores are provided for each dataset listed in the *Data* column, separated by commas (Example: score$_{Data1}$, score$_{Data2}$). When separate results are provided for different categories in a dataset (e.g., *True, False*), these categories are shown in brackets in the *Data* column. The same structure is used in the *Evaluation* column to report the specific results for each category (Example: score$_{True}$, score$_{False}$).

Methods	Pub and baseline	Data	Evaluation	Ablation: No aff. feats
Methods Combining Emotion with Other Text-Based Features	FakeFlow [122]	MultiSourceFake, LUN	F1-macro 0.96, 0.96	MultiSourceFake 0.91
	Baseline (BERT)	MultiSourceFake	F1-macro 0.93	
	Mixture-of-Experts[178]	Weibo21	F1 0.9223	0.9185
	Baseline (BERT)		F1 0.8795	
	EmoAttentionBERT[164]	MultiFC (snopes, politifact)	F1-macro 0.344, 0.318	
	Baseline (BERT)		F1-macro 0.295, 0.282	
	LSTM (with fuzzy sentiment) [165]	Combination of Infodemic and CoAID	F1 0.9143	0.9024
Mining of Dual Emotions	AGWu-RF [20]	RumourEval19, PHEME, Fakeddit	F1 0.95, 0.97, 0.97	
	Baseline(RumorEval19,PHEME:DTCA[213]; Fakeddit:DeepNet[214])		F1 0.82, 0.83, 0.83	
	MDE [29]	RumourEval19, Weibo16, Weibo20	F1-macro 0.346, 0.867, 0.915	
	EFN [135]	Customized	F1 0.874	0.859
	Baseline (GRU)		F1 0.84	
	[173]	Fakeddit	F1 0.9781	0.9554
Methods Based on Tree or Graph Structures	SSE-BERT [23]	Weibo16, Weibo20	F1 0.947, 0.943	0.941, 0.94
	Baseline (Bi-GCN)		F1 0.892, 0.882	
	ptVAE [175]	Weibo16, Twitter15, Twitter16 (True, False)	F1 (0.853,0.848), (0.67,0.697), (0.682,0.682)	(0.776,0.754), (0.6,0.638), (0.641,0.676)
	Baseline (GFVAE [215])		F1(0.752,0.745),(0.623,0.653),(0.639,0.648)	
	[179]	Weibo16	F1 0.97	
	Baseline (BERT)		F1 0.88	
	MHN [176]	LUN, SLN	F1-macro 0.7169, 0.8972	(LUN) no sentiment net 0.6983
	Baseline (GCN+Attn)		F1-macro 0.6642, 0.8524	
	SA-HyperGAT [163]	Twitter15, Twitter16 (UR, NR, TR, FR)	F1 (0.857,0.838,0.923,0.88), (0.925,0.886,0.957,0.86)	(0.837,0.763,0.905,0.861), (0.866,0.765,0.939,0.87)
	Baseline (Bi-GCN)		F1 (0.752,0.772,0.885,0.847), (0.818,0.772,0.885,0.847)	
	GCS [166]	PHEME	F1 0.9342	
	Baseline ([153]GB)		F1 0.8496	
Methods Based on Temporal Information	TDEI [22]	Weibo16 (True, False)	F1 (0.969,0.968)	(0.959,0.958)
	Baseline(RvNN)		F1 (0.911,0.905)	
	SD-TsDTS-CGRU [154, 155]	Weibo16, Twitter16-2 (True, False)	F1 (0.963,0.963), (0.880,0.889)	Weibo16(rumor) 0.92
	Baseline(GRU)		F1 (0.830,0.835), (0.796,0.804)	
	RvNNwithTemporal[170]	Weibo16, RumourEval19	F1-macro 0.939, 0.534	0.925,0.492
	Baseline(RvNN)		F1-macro 0.919, 0.506	
Multitask Learning	[30]	PHEME, FakeNews AMT, Celeb, Gossipcop	F1 0.864, 0.866, 0.879, 0.778	0.848,0.806,0.815,0.745
	[182, 183]	Source: FakeNewsAMT; Target: Gossipcop (cross domain)	Accuracy 0.795	0.451
	[161]	ByteDance, FNC, Covid-Stance	F1 0.9974, 0.9688, 0.9859	0.8821, 6826, 0.8428
	Baseline(SiameseLSTM[216])		F1 0.8783, 0.675, 0.8392	
Multimodal Methods	[174]	PolitiFact, GossipCop	F1 0.92, 0.894	0.914, 0.892
	Baseline (BERT)		F1 0.818, 0.850	
	[151]	PolitiFact, GossipCop	F1-macro 0.7724, 0.8042	0.7085, 0.7091
	Baseline (SVM)		F1-macro 0.6557, 0.6124	
	[133]	MMM (Real, Fake)	F1 (0.960,0.949)	0.926, 0.907
	Baseline(MLBERT+ResNet)		F1 (0.765,0.703)	
	[132]	NovEmoFake	F1-micro 0.9775	0.9054
	Baseline(BERT,ResNet)		F1-micro (0.8002, 0.7401)	
	[168]	Fakeddit	F1 0.9329 Accuracy 0.9194	
	Baseline(BERT+ResNet)		Accuracy 0.8909	
Machine Learning	[152] (RF)	LIAR	Accuracy 0.356	
	[153] (SVM)	PHEME	F1 0.86	
	[157] (GB)	Fake.Br	Accuracy 0.9253	
	[162] (AB)	FakeNewsSet	Accuracy 0.95	
	[136] (GB)	Portuguese(customized),Fake.br	F1 0.832, 0.949	
	[158] (DT)	PolitiFact, GossipCop	Accuracy 0.76, 0.74	
	[160] (LR)	PolitiFact	F1 0.61	0.33
	[24] (RF)	ISOT	F1-macro 0.868	

Emotion detection for misinformation

Methods	Pub and baseline	Data	Evaluation	Ablation: No aff. feats
Machine Learning	[21] (AdaBoost)	[102]	F1 0.8404	
	[169] (SVM)	LIAR-PLUS (True, False)	F1 (0.77, 0.80)	
	[171] (RF)	Buzzfeed Political News	F1 0.949	
	[125] (AdaBoost)	Customized	F1 0.9343	
	[177] (RF)	Twitter_harvard, Health_related	F1 0.77, 0.8509	
	[177](RF+BERTfeature)	Twitter_harvard, Health_related	F1 0.9664, 0.8693	
Deep Learning	[153](LSTM-HAN)	PHEME	F1 0.84	
	[28] (MARBERT)	Customized	F1 0.8886	0.8764
	[160] (DNN)	PolitiFact	F1 0.64	0.32
	[21] (GRU)	[102] (Real,Fake)	F1 0.88, 0.83	
RDES	[43]	RumourEval17	Accuracy 0.464	
	[204]	RumourEval17	Accuracy 0.53	
	[205]	RumourEval17	0.422	
	[209]	RumourEval19	F-macro 0.262	
	[42]	RumourEval19	F-macro 0.5868	
	[140]	DAST	F1 0.68	
LLMs	[217] GPT-4	100 fact-checked news items	Accuracy 0.71	
	[130] ChatGPT	Twitter15,PHEME,FND23, FakeNewsNet,LTCR	Accuracy 0.4286,0.5513,0.8895,0.8055,0.69	
	[130] LLaMA2-7B	Twitter15, PHEME,FND23, FakeNewsNet,LTCR	Accuracy 0.2798,0.516,0.5436,0.6898,0.6026	
	[25] ChatGPT	Weibo21, GossipCop	macro-F1 0.725, 0.702	
	[25] LLM+SLM	Weibo21, GossipCop	macro-F1 0.790, 0.801	
	[27] FactLLaMA-7B	LIAR,RAWFC	F1 0.3044,0.5565	
	[218] ChatGPT-Assisted Network	Twitter15, Twitter16 (UR, NR, TR, FR)	F1 (0.901,0.816,0.870,0.811),(0.862,0.838,0.802,0.826)	

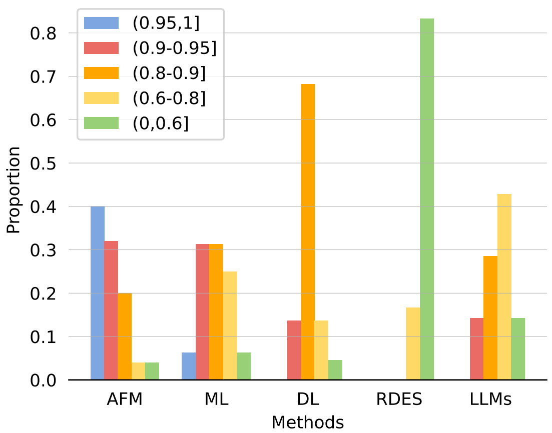


Figure 12: Distributions of evaluation scores obtained across different method categories. AFM: Advanced Fusion Methods. RDES: Rumor Detection based on Emotion and Stance

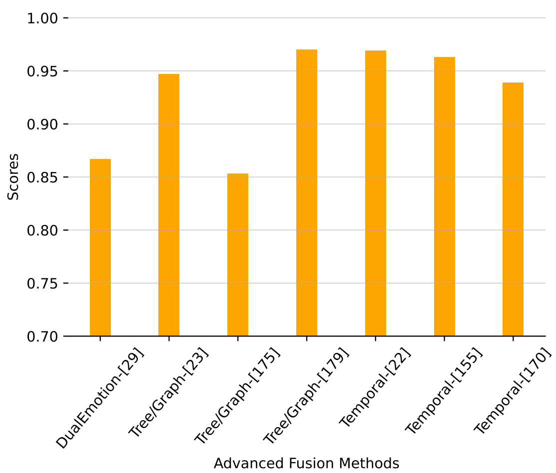


Figure 13: Performance of different methods evaluated on the Weibo16 dataset

the advanced fusion methods are the same. However, the performance of the baselines is usually noticeably inferior to that of advanced fusion methods. This demonstrates that the more complex processing strategies of advanced fusion methods are able to make more effective use of features relating to emotions and other aspects of misinformation, resulting in superior performance.

Comparison between different fusion methods: Although the experimental results for EmoAttention BERT [164] highlight the importance of emotionally charged style, and in particular of emotional intensity, as a predictive feature of fake news, Table 5 illustrates that the performance of this method is low. A probable reason is that the method is evaluated on a complex dataset that includes multiple domains and labels, but the fairly simple framework fails to account for potential differences in the characteristics of data across different domains. A possible solution is to adopt a multitask architecture, similar to [30, 182, 183], in which a domain classifier is incorporated as a discriminator to ensure that the model performs well across multiple domains, through reinforcement learning. Similarly, the challenging characteristics of the RumourEval19 dataset (i.e., low inter-annotator agreement and sparse data [142, 219]) result in relatively low performance by the RvNN with Temporal

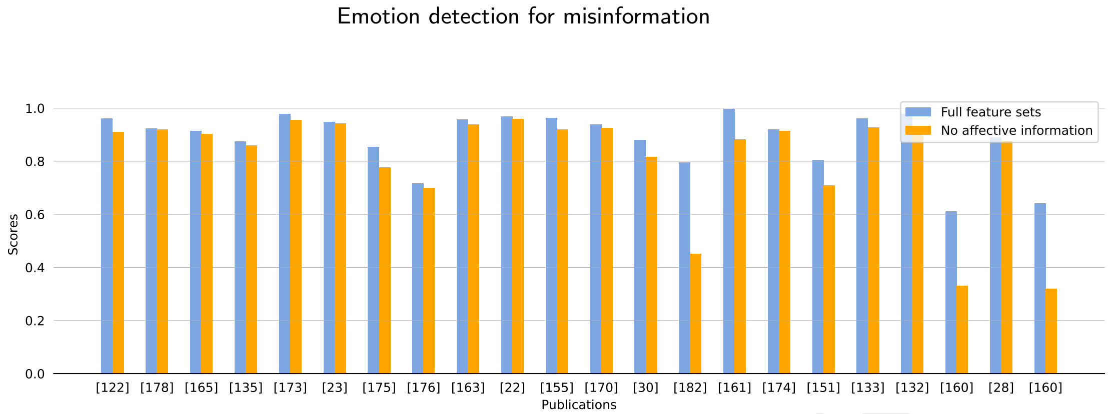


Figure 14: Comparison of full feature sets with affective information ablation.

[170] (0.53 F1) and MDE [29] (0.35 F1) methods, even though they achieve much higher results when applied to other datasets. In comparison, AGWu-RF [20] attains vastly superior results on RumourEval19 (0.95 F1). In common with MDE, AGWu-RF uses dual emotion features. However, its combination of these features with an RF with genetically adapted weights appears to make it robust in handling this problematic dataset. Furthermore, AGWu-RF is demonstrated to be sufficiently generalizable for successful application to social media datasets with varying characteristics, e.g., it outperforms both [30] and [166] on the PHEME dataset. It is also notable that while AGWu-RF uses only textual information, it achieves better results than [168] on the multimodal Fakeddit dataset, even though the latter method uses both text and image features.

Source tweets in Twitter15 and Twitter16 are annotated with four class labels, i.e., non-rumor (NR), false rumor (FR), true rumor (TR), and unverified rumor (UR). While these datasets are used to evaluate both the ptVAE [175] and SA-HyperGAT [163] methods, the evaluation of ptVAE uses only two classes, i.e., *true* (non-rumors and true rumors) and *false* (false rumors). Nevertheless, ptVAE exhibits lower performance than SA-HyperGAT on these datasets, and its performance on the Weibo16 dataset is also inferior to several other methods. This could be due to the less rigorous data processing methods used in ptVAE, but it is more likely that the proposed VAE architecture for sentiment analysis is not as effective as other methods, such as the fine-tuned RoBERTa model used in SA-HyperGAT [163]. Table 5 shows that the Weibo16 dataset is used to evaluate a large number of methods; Figure 13 provides a visual comparison of the performance of several of these methods on this dataset. The figure provides strong evidence that combining sentiment/emotion features with those accounting for propagational and/or temporal features is highly important. Specifically, [179], TDEI [22], SD-TsDTS-CGRU [154, 155], and RvNN with Temporal [170] achieve the highest levels of performance on Weibo16 (0.94 F1 or higher). On the same dataset, the impressive F1 of 0.95 achieved by SSE-BERT [23], which combines sentiment and dependency tree information from source posts, indicates the potential value of considering syntactic information. Meanwhile, MDE [29] and ptVAE [175] exhibit the lowest performance, with scores below 0.9. Both FakeFlow [122] and MHN [176] analyze changes in affective information across the different parts of article, and were evaluated on the LUN dataset. However, the superior performance of FakeFlow suggests that accounting for affective interactions with different topics represents a more successful approach.

Feature Fusion: Textual data contains an abundance of information, which has been encoded using a wide variety of features in different misinformation detection methods, as illustrated in Figure 4. In Table 3, we list the specific features that have been combined with emotion and/or sentiment information in different studies. A comparison of the performance of complete models with those of ablation experiments in Table 5 and Figure 14 confirms the importance of sentiment and emotion features in misinformation detection, since the removal of these features always results in a drop in performance, sometimes by a significant margin. However, high levels of performance can only be achieved by combining multiple features. For example, it is shown in [122] that the proposed combination of topic and affective features outperforms the use of either topic or affective features in isolation. Furthermore, [132, 133] show that extracting features from the images that accompany posts on social media platforms can provide additional clues about the emotional states and behaviors of individuals, and thus helps to boost the results of misinformation detection.

Model Fusion: Different models and learning techniques possess their own strengths and weaknesses, and optimal misinformation detection performance can generally only be achieved by combining a number of different techniques. For example, pre-trained models like GloVe and BERT are effective in encoding textual content with word embeddings, while RoBERTa can be successfully employed for sentiment and emotion detection [220]. Meanwhile, methods like CNN, LSTM, or GRU may be usefully adopted for feature extraction. Encoding information about the graph structure inherent in many datasets requires different approaches. For instance, dependency and sentiment trees may be used to represent grammatical or semantic aspects of sentence structure, while GCN and hypergraphs can encode the tree-like structure of social media data. To capture temporal features of rumor propagation, different studies have utilized RNN, LSTM, and GRU models, which excel in handling time series data. The application of fusion or ensemble techniques can fully leverage the relative strengths of these different types of methods and models, as may be confirmed by comparing the results of the advanced methods with baseline methods in Table 5.

The analysis above underlines the complexities of developing effective misinformation detection models. High levels of performance can only be achieved by leveraging multiple relevant features, which include thematic, temporal, propagation structure, dual emotion, and/or image information, in addition to sentiment and emotion. Furthermore, the specific methods chosen to learn or represent these features can also impact upon performance, and it is usually necessary to combine a range of learning methods to achieve optimal results. Moreover, to ensure cross-domain robustness, the employment of multitask learning frameworks incorporating reinforcement learning can be advantageous.

5. Challenges and Future Research Directions

While this article has reviewed a large and diverse body of research relating to emotion-based misinformation detection, there still remain a variety of unsolved challenges in the field. In this section, we outline the most important of these challenges, and discuss potential future directions of research.

5.1. Dataset Collection (Multi-platform, Multilingual)

There are many popular social media platforms, such as Twitter, Facebook, Reddit, and Sina Weibo, among others, which constitute major means of spreading misinformation. These platforms use different languages and varied data formats, which can make the processing of such data cumbersome. Given that the dissemination of fake news is a global problem, it is important to develop approaches that are more universally applicable than most currently available methods. However, achieving this goal is hindered by the limitations of the majority of currently publicly available misinformation datasets, which are usually collected from a single platform (as shown in Table 2) and which predominantly concern textual data in a single language (typically English or Chinese). Only by developing larger and more diverse datasets will it become possible to develop more general models, which are urgently needed. These datasets should cover multiple data formats, obtained from different platforms, and should include multiple languages.

5.2. Annotation (Emotion)

The development of emotion-based misinformation detection methods with optimal performance requires that supporting misinformation datasets are annotated with reliable emotion and/or sentiment labels, since inaccurate labels are likely to impact negatively on the overall performance of the methods. While such labels are most often obtained using dictionary lookup, some studies have employed transfer learning methods, by applying models trained on other sentiment analysis or emotion-labeled datasets, to automatically annotate the emotions expressed in misinformation datasets. Examples include [30, 182, 183], which use a previously developed Unison model, and [163], which utilizes a fine-tuned RoBERTa model. However, the emotion labels obtained by applying these methods are not sufficiently accurate. Compared to time-consuming manual annotation, a more promising approach is to use LLMs to annotate emotion and/or sentiment [74, 75, 76], given their advanced capabilities and transferability. Both [75] and [76] have demonstrated that LLMs can compete with or exceed the SOTA in recognizing emotions in dialogue. In particular, [75] shows that the LLaMA-7B [221] model can achieve performance levels close to those of SOTA supervised methods, but using only half as much training data for fine-tuning. Zhang et al. [74] propose an instruction-tuned LLM for financial sentiment analysis which, augmented with additional context from external sources, is able to outperform LLM baselines such as ChatGPT and LLaMA by margins of between 15% and 48%. The above studies all highlight the tremendous potential of LLMs in the field of sentiment analysis.

5.3. Multimodality

Although rumors and fake news were traditionally spread through face-to-face communication, the emergence of social media resulted in their primary means of dissemination switching to text. However, continual advances in technology have led to an increasing shift towards multimodality. For example, people now frequently augment textual post content with images or videos, while on platforms like YouTube or TikTok, videos are the predominant means of sharing information. Accordingly, it is becoming increasingly important to explore methods that can address the challenges of multimodality [222], and that are able to adapt to the ever-changing characteristics of social media communication. While we have reviewed a number of approaches that combine text and image-based information, recent advanced multimodal models that integrate language and visual understanding provide considerable scope for further research in this area. For example, GPT-4 [9] has a certain level of visual understanding capability, although its implementation details have not been publicly disclosed. Inspired by the success of LLMs, some studies have started to focus on large multimodal models, such as LLaVA [223, 224], which is an end-to-end large multimodal model that connects a visual encoder with a large language model to achieve general visual and language understanding. Additionally, MiniGPT-5 [225] introduces a novel interleaved vision-and-language generation technique, with a focus on non-descriptive multimodal generation. Exploring the integration of these large multimodal models within misinformation detection methods constitutes an interesting and promising research direction.

5.4. Benchmarks

Several benchmark datasets have been developed in the context of shared tasks, which are aimed at evaluating various different characteristics of misinformation detection methods. For example, the datasets created for RumourEval17 [33] and RumourEval19 [142] focus on stance and misinformation detection, while the CLEF2020 - CheckThat! Lab [96] addresses multilinguality through the inclusion of benchmark data in both English and Arabic. These datasets are complemented by a recently developed multimodal benchmark for fake news detection [226]. Despite the value of these datasets in facilitating the evaluation of various different *individual* aspects of methods, there is still a lack of a suitably comprehensive benchmark that can *simultaneously* evaluate the ability of misinformation detection methods to handle diverse types of multimodal data from multiple platforms that cover different languages, as well as assessing their ability to perform important subsidiary tasks, such as emotion, sentiment and stance detection, identification of rumor source, etc. We believe that the development of such a dataset would be of enormous value in helping to guide research in this area towards the development of more robust and universally applicable misinformation methods, as well as focusing attention on the development of important supporting technologies.

5.5. Interpretability

Understanding how and why misinformation detection models have arrived at their decision about whether or not a post or news article represents true or fake information can be important to make their reasoning processes more transparent, and to make it easier to understand why errors occur. However, despite the high levels of performance achieved by many DL approaches, their black-box nature means that no such reasoning information is available, and that their decisions are hard to justify. Although it remains a challenge to develop models whose results are both sufficiently accurate *and* interpretable, several studies have proposed possible solutions for *explainable* misinformation detection. These include the use of topic-based features for classification [227], Explainable Artificial Intelligence (XAI) techniques [228], and Commonsense Knowledge Graphs [229]. Recent research has also begun to focus on the development of interpretable LLMs [230], such as MentalLLaMA [231], which is an interpretable mental health analysis model based on LLaMA2. Accordingly, it is hoped that researchers working on misinformation detection will begin to place greater emphasis on exploring the increasing range of options that could be used to improve the interpretability of their models.

5.6. Large Language Models

The popularity of ChatGPT and GPT-4 [9] has resulted in the powerful capabilities of LLMs becoming widely known [232]. As mentioned above, there is potential for LLMs to be employed in misinformation detection in multiple ways, including sentiment and emotion detection, multimodal analysis, and to enhance the interpretability of detection models. Some studies have additionally begun to explore the use of LLMs for rumor and fake news prediction. For example, Hu et al. [25] propose a framework for fake news detection in which a small language model (i.e., BERT) is complemented by an LLM, which provides multi-perspective guiding principles to improve prediction accuracy. Meanwhile, Pavlyshenko et al. [26] design prompts to fine-tune LLaMA2 for rumor and fake news detection. Cheung et al. [27] use external knowledge to bridge the gap between knowledge encoded in an LLM and the most up-to-date information available on the Internet, in order to enhance fake news detection performance. Although exploring the capabilities of LLMs in the context of misinformation detection is currently still in its infancy, the promising results achieved by these initial approaches, combined with the indisputable power and advanced capabilities of LLMs, motivate further exploration of how they can be best exploited to improve the accuracy of rumor and fake news detection.

6. Threats to Validity

In Section 4.6, we presented the performance of a range of emotion-based misinformation detection methods and discussed the strengths and weaknesses of different approaches. To facilitate ease of comparison, we strived to select methods that were evaluated using common datasets. Nevertheless, there remain a number of factors that could impact upon the validity of the comparisons made in the discussion, and thus represent a potential limitation of this review. These factors are discussed below.

Emotion detection accuracy: Different misinformation detection methods use a range of different techniques to detect information about emotions and/or sentiments. The most commonly used of these are presented in Table 3, encompassing both dictionary-based methods (e.g. VADER and TextBlob) and transfer learning methods. The latter methods involve fine-tuning models on sentiment analysis datasets and then using these fine-tuned models to automatically detect affective information in fake news datasets. The accuracy of the recognized emotion/sentiment information has a direct impact on the performance of misinformation detection. Although dictionary-based methods are simple and convenient to use, their performance can be low and unstable [233]. The generalization of transfer learning methods also requires further investigation. These issues may impact upon the outcomes of our comparative discussion of methods in Section 4.6. With the exception of manual annotation, there is currently a lack of techniques for analyzing affective information that exhibit robust performance. As discussed in Section 5.2, one promising way forward may be the development of emotion analysis models based on LLMs.

Feature selection: The use of different feature sets can also affect the results of misinformation detection. As shown in Table 3, the features employed in different studies vary significantly, both in terms of affective information, such as sentiment or varying numbers of emotions, and other types of features. For example, methods evaluated using the Twitter16 dataset include [155], which utilizes a combination of sentiment, emotion and temporal features; [163], which only employs sentiment information and structural information; and [175], which incorporates user-based, structural, and propagation features, in addition to sentiment information.

Dataset processing: Even if multiple methods use the same dataset for training and/or evaluation, the application of different processing techniques to the dataset could potentially affect the results reported. For example, while some datasets are pre-divided into training and test sets, others may be randomly split in different ways by different researchers. This random partitioning can affect the comparison between alternative methods. Moreover, datasets with complex formats, such as data from Twitter, may or may not undergo pre-processing prior to use by different methods. For example, emojis and hashtags can cause issues for some text processing tools, and hence some studies pre-process the tweets to remove them. On the other hand, emojis and hashtags can convey valuable information, and so some studies preserve them.

Base model selection: Different base models have different characteristics, which can lead to varying outcomes when they are applied to the task of misinformation detection. For example, among the advanced fusion methods that focus on the use of temporal information, each uses a different type or combination of base models, i.e., TDEI utilizes GRU and BERT, SD-TsDTS-CGRU employs cascaded GRU, while [170] uses RvNN.

By considering the above factors, future endeavors to address emotion-based misinformation detection will be better equipped to develop more optimal solutions. According to the previously demonstrated vital role of emotion information in misinformation detection, it is of central importance to develop more robust and accurate emotion detection methods. Moreover, to determine the most suitable methods for different types of datasets, it is crucial that more exhaustive studies are carried out to determine the best combinations of features and base models, based on a dataset that has been split in a particular way and that has undergone a specific pre-processing strategy.

7. Conclusion

The unstoppable growth of social media is making it easier than ever for misinformation to spread rapidly and widely. As such, there is an increasingly urgent need for robust automated methods that can detect and stop this spread as efficiently and effectively as possible. In this article, we have comprehensively analyzed approaches to rumor and fake news detection based on emotions and/or sentiment, with a focus on advanced fusion methods. After introducing related work, we firstly motivated these approaches by summarizing research that confirms the strong links between emotion and misinformation. We subsequently provided an overview of available datasets that can support the development of misinformation detection methods, followed by a summary of both conventional ML and DL methods that have been employed in emotion-based misinformation detection approaches. We then proceeded to describe and categorize a diverse range of recently proposed advanced fusion methods that combine the use of emotion/sentiment with various other features, and/or integrate a number of different learning methods to achieve their goals. We additionally provided an overview of emotion-based stance detection methods in misinformation. Subsequently, we reported the performance of the advanced fusion methods, compared them to the results achieved by simpler approaches, and discussed the relative strengths and weaknesses of different advanced methods from various perspectives.

We outlined six important types of unresolved challenges in the field of rumor and fake news detection. For each type of challenge, we provided suggestions of possible future research that could help to address and resolve these challenges. Firstly, to develop methods that better account for the global-level spread of misinformation across multiple social media platforms, it is important to construct more wide-coverage misinformation datasets, whose data is multilingual and extracted from a variety of different platforms with varying data formats. Secondly, to take better advantage of affective information expressed in the context of misinformation, it is important that greater emphasis is placed on developing robust, accurate and reliable emotion detection methods. Thirdly, the rapidly changing face of social media data means that there is an urgent need to develop misinformation detection approaches that effectively extract, combine and interpret features obtained from multiple modes of communication, including text, images and video. Fourthly, the effective evaluation of such complex methods and their constituent tasks can only be achieved through the development of comprehensive novel benchmarks, whose data is both multimodal and multilingual, and which include multiple levels of annotation that allow the performance of important subsidiary tasks, such as emotion and stance detection, to be effectively assessed. Fifthly, understanding the impact of different types of reasoning on misinformation detection performance is dependent on a switch from traditional DL techniques to more explainable methods, which will make it easier to understand why errors are occurring and how best to handle them. Finally, developing a better understanding of the role that LLMs can play in the detection of misinformation is dependent on further investigation to determine how best to exploit the potential of LLMs within this field.

Additionally, we discussed several threats to the validity of our comparative analysis of different methods, including the generally unstable performance of emotion detection and sentiment analysis tools, along with the potential impacts of using varying sets of features and different dataset processing methods, and of selecting alternative base models. These are all important factors that should be considered in future efforts to develop more effective emotion-based misinformation detection methods.

In summary, our review has demonstrated the significant role of sentiment and emotion in misinformation, and has aimed to highlight the most important aspects in its automated detection. It is intended that the survey will enable researchers who are interested in the field to better appreciate the tremendous value of affective information in misinformation detection, and will help to drive further advances to the SOTA in this field.

Acknowledgments

This work has been supported by the scholar award from the Department of Computer Science at the University of Manchester.

A. Abbreviations and Notations (Table 6)

B. Specific Types of Content-based Features (Table 7)

C. Emotion Detection Tools (Table 8)

Various tools and resources have been used for the detection of emotion and sentiment features, the most commonly used of which are listed and briefly described in Table 8. Various other tools are also available, including the following: Emojis Dictionary[9], Emoticons list[10], Affect-Br[248], SemEval[11], MPQA[12], ENGAR[249], Hespress Facebook[13], Offense lexicon[14], Sarcasm lexicon[250], Named entities lexicon (Religion lexicon, Nationality lexicon, Named entities)[159], Baidu sentiment API[15], NLTK sentiment module[16], SEO Scout's analysis tool[17], IBM Watson's Natural Language Understanding (NLU)[18], MeaningCloud[19], ParallelDots[20], Empath[251], EffectWordNet[252], Hu&Liu opinion lexicon[21], SSWE[253], NRC-Canada[254], Stanford sentiment Tree[255], Dictionary of Affect in Language (DAL) [256], Affective Norms for English Words (ANEW) [247], MetaMind sentimentclassifier API[22].

[9]https://drive.google.com/file/d/1G1vIkkbqPBYPKHcQ8qy0G2zkoab2Qv4v/view

[10]https://en.wikipedia.org/wiki/List_of_emoticons

[11]http://www.saifmohammad.com/WebPages/SCL.html

[12]http://www.purl.org/net/ArabicSA

[13]https://fr-fr.facebook.com/Hespress

[14]https://sites.google.com/site/offensevalsharedtask/

[15]https://ai.baidu.com/tech/nlp_apply/sentiment_classify

[16]https://www.nltk.org/api/nltk.sentiment.html?highlight=sentiment#module-nltk.sentiment

[17]https://seoscout.com

[18]https://https://www.sciencedirect.com/topics/computer-science/natural-language-understanding

[19]https://www.meaningcloud.com/

[20]https://apis.paralleldots.com/text_docs/index.html

[21]http://www.cs.uic.edu/liub/FBS

[22]https://www.metamind.io

D. Evaluation Measurements

D.1. Misinformation Detection Evaluation Measurements

A variety of techniques has been used to evaluate the output of misinformation detection methods, including accuracy, recall, precision, F1-score, F1-macro, class-wise F1-score, AUC [173, 177, 178], and RMSE [29]. These are calculated on the basis of a number of basic concepts, which are defined as follows: TP (True Positive) refers to the number of samples that the model correctly predicts as positive; TN (True Negative) refers to the number of samples that the model correctly predicts as negative; FP (False Positive) refers to the number of samples that the model incorrectly predicts as positive; FN (False Negative) refers to the number of samples that the model incorrectly predicts as negative.

Accuracy indicates the overall classification correctness of a model:

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

Recall measures the model's ability to identify positive-class samples:

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

Precision measures the proportion of true positive samples among the samples predicted as positive by the model:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

The F1 score represents the harmonic mean of precision and recall:

$$F1_{score} = \frac{2 * Recall * Precision}{Recall + Precision} \tag{4}$$

F1-macro is used to evaluate the performance of a multi-class classifier, by combining the $F1_{score}$ of each class; Class-wise $F1_{score}$ refers to the $F1_{score}$ of each individual class, and can be used to evaluate the performance of the classifier for each class. The AUC (Area Under the Curve) is a commonly used metric for evaluating the performance of classification models. It measures the predictive ability of the model by calculating the area under the ROC (Receiver Operating Characteristic) curve.

Root Mean Squared Error (RMSE) represents the expected value of the squared error.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|^2} \tag{5}$$

Table 6
Abbreviations and notations used in the article

Notation	Explanation	Notation	Explanation	Notation	Explanation
LLMs	Large Language Models	HAN	Hierarchical Attention Networks	SPM	Sentiment Pattern Module
AI	Artificial Intelligence	GNN	Graph Neural Network	PFM	Propagation Feature Module
DL	Deep Learning	AF	Affective Features	MHN	Mixed Heterogeneous Network
ML	Machine Learning	E	Emotion	SA-HyperGAT	Sentiment-Aware Hypergraph Attention Network
CNN	Convolutional Neural Networks	S	Sentiment	GCS	Graph-based Contextual and Semantic learning method
RNN	Recurrent Neural Networks	IE	Image Emotion	TDEI	Temporal Dynamic structure and Emotional Information
GCN	Graph Convolutional Networks	ERD	Emotion-based Rumor Detection	SAME	Sentiment-Aware Multimodal Embedding for detecting fake news
ED	Emotion Detection	MLs	Various Machine Learning methods	SCL	Supervised Contrastive Learning
SA	Sentiment Analysis	EmoLex	NRC Emotion Lexicon	ELA	Error Level Analysis
PAD	Pleasure-Arousal-Dominance	ALO	Affective Lexicon Ontology	DAST	Danish stance-annotated dataset
VAD	Valence-Arousal-Dominance	VADER	Valence Aware Dictionary for sEntiment Reasoning	ESD	Emotion-based Stance Detection
RAM	Relationship Analysis Methods	EmoSN	EmoSenticNet	RDES	Rumor Detection based on Emotion and Stance
PANAS	Positive-Negative Emotional Scale	ANEW	Affective Norms for English Words	SLM	Small Language Model
LR	Logistic Regression	e_a	sentence-level Embeddings from attention	AFM	Advanced Fusion Methods
MANOVA	Multivariate Analysis of Variance	e_s	sentiment Embeddings	NR	Non-Rumor
MANCOVA	Multivariate Analysis of Covariance	e_d	domain Embeddings	FR	False Rumor
ANOVA	Analysis of Variance	C_i	Collaborative influence function	TR	True Rumor
ANCOVA	Analysis of Covariance	MDE	a method based on Mining Dual Emotion for fake news detection	UR	Unverified Rumor
A	Available	emo_S	Social emotion	XAI	eXplainable Artificial Intelligence
N	No link	emo_{gap}	Emotion gap	BoW	Bag of Words
R	Request	emo_P	Publisher emotion	NLTK	Natural Language ToolKit
NB	Naive Bayes	emo_S^{mean}	average social emotion	TP	True Positive
KNN	k-Nearest Neighbour	emo_S^{max}	extreme social emotional	FR	True Negative
SVM	Support Vector Machine	MLP	Multi-Layer Perceptron	FP	False Positive
RF	Random Forest	EFN	Emotion-based Fake News detection framework	FN	False Negative
DT	Decision Tree	Gate_N, Gate_C, Gate_M	different Gate recurrent units	AUC	Area Under the Curve
AB	AdaBoost	SSE-BERT	Syntax and Sentiment Enhanced version BERT	ROC	Receiver Operating Characteristic
GB	Gradient Boost	SOTA	State-Of-The-Art	RMSE	Root Mean Squared Error
DNN	Deep Neural Network	PLMs	Pre-trained Language Models		

D.2. Stance Detection Evaluation Measurements

Accuracy, recall, precision, F1-score, F1-macro, class-wise F1-score, FNC1-Score [206, 207], and weighted accuracy [211] have all been used to evaluate stance detection methods.

The FNC1 weighted accuracy score is used as the final evaluation metric for the FNC dataset, and is calculated as follows:

$$FNC1_{score} = 0.25 * Accuracy_{Unrelated} + 0.75 * Accuracy_{Agree,Disagree,Discuss} \tag{6}$$

Weighted Accuracy is a performance metric that takes into account the weight of each class in an imbalanced dataset. It calculates the overall performance of the model by using the weighted average of the accuracy for each category.

CRediT authorship contribution statement

Zhiwei Liu: Writing – original draft, conceptualization, methodology, data curation, visualization, review & editing. **Tianlin Zhang:** Writing – original draft, review & editing. **Kailai Yang:** Writing – original draft, review & editing.

Table 7
Specific types of content-based features

Term	Features
Similarity Features	title-text similarity, word similarity, sentence similarity, cosine similarity between source post and related comments
Cluster Features	word-cluster feature, brown cluster feature [203], SDQC depth-based clusters [203]
Semantic Feature	word vector features (GloVe [234], BERT [235], GoogleW2V [236], Word2Vec [195])
Grammatical Features	part-of-speech tags, noun, verbs, adjectives, and pronouns
Lexical Features	bad sexual words, cue words, multilingual hate lexicon, linguistic words, specific categories, denial term, support words, negation words, swear words, surprise and doubt words
Linguistic-informed Features	tf-idf, n-gram, named entity recognition, text language, bag-of-characters, bag of words (BoW)
Stylistic Features[41]	question marks, exclamation marks, punctuation marks, sentence length, uppercase ratio, consecutive characters and letters, presence of URLs, number of stop words, number of upper case letters, number of lower case letters, number of numeric values, word count, character count, sentence count, average sentence length, ease of comprehension, lexical diversity
Syntactic Features	ratio of negation, bag of relations (all tokens, list of words, verbs)
Conversation based Features	text similarity to source tweet, text similarity to replied tweet, tweet depth
Twitter Metadata [28, 208]	number of characters in a tweet, number of retweets, favorites, presence of hashtags, URLs, mentions, existence of photos, creation time gaps for posts, Twitter verification. etc.
Reddit Metadata [140]	karma, gold status, Reddit employment status (if any), verified e-mail, reply count, upvotes, whether the user is the submission submitter. Reddit commenting syntax: sarcasm ('/s'), edited ('edit:'), and quote count ('>')
Others	topics, term features, textual novelty

Paul Thompson: Writing – original draft, review & editing. **Zeping Yu:** Writing – review & editing. **Sophia Ananiadou:** Writing – review & editing.

References

[1] M. Fernandez, H. Alani, Online misinformation: Challenges and future directions, in: Companion Proceedings of the The Web Conference 2018, 2018, pp. 595–602.

[2] H. Allcott, M. Gentzkow, Social media and fake news in the 2016 election, Journal of economic perspectives 31 (2) (2017) 211–236.

[3] P. Bordia, N. DiFonzo, Rumor, gossip and urban legends, Diogenes 54 (1) (2007) 19–35.

[4] T. A. Maniou, Semantic analysis of cultural heritage news propagation in social media: Assessing the role of media and journalists in the era of big data, Sustainability 13 (1) (2021) 341.

[5] D. A. Scheufele, N. M. Krause, Science audiences, misinformation, and fake news, Proceedings of the National Academy of Sciences 116 (16) (2019) 7662–7669.

[6] R. Li, Y. Wang, et al., Research on the spread and governance of internet rumors under the covid-19, Academic Journal of Humanities & Social Sciences 4 (7) (2021) 38–42.

[7] M. Cheng, S. Wang, X. Yan, T. Yang, W. Wang, Z. Huang, X. Xiao, S. Nazarian, P. Bogdan, A covid-19 rumor dataset, Frontiers in Psychology 12 (2021) 644801.

[8] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, et al., Training language models to follow instructions with human feedback, Advances in Neural Information Processing Systems 35 (2022) 27730–27744.

[9] R. OpenAI, Gpt-4 technical report, arXiv (2023) 2303–08774.

[10] Y. Yan, B. Li, J. Feng, Y. Du, Z. Lu, M. Huang, Y. Li, Research on the impact of trends related to chatgpt, Procedia Computer Science 221 (2023) 1284–1291.

[11] R. N. Zaeem, C. Li, K. S. Barber, On sentiment of online fake news, in: 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 2020, pp. 760–767.

[12] N. Zhang, J. Song, K. Chen, S. Jia, Emotional contagion in the propagation of online rumors., Issues in Information Systems 23 (2) (2022).

[13] C. G. Horner, D. Galletta, J. Crawford, A. Shirsat, Emotions: The unexplored fuel of fake news on social media, Journal of Management Information Systems 38 (4) (2021) 1039–1066.

[14] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, science 359 (6380) (2018) 1146–1151.

[15] N. Pröllochs, D. Bär, S. Feuerriegel, Emotions in online rumor diffusion, EPJ Data Science 10 (1) (2021) 51.

[16] N. Pröllochs, D. Bär, S. Feuerriegel, Emotions explain differences in the diffusion of true vs. false social media rumors, Scientific Reports 11 (1) (2021) 22721.

[17] W. Dong, J. Tao, X. Xia, L. Ye, H. Xu, P. Jiang, Y. Liu, Public emotions and rumors spread during the covid-19 epidemic in china: web-based correlation study, Journal of Medical Internet Research 22 (11) (2020) e21933.

[18] J. Cui, Z. Wang, S.-B. Ho, E. Cambria, Survey on sentiment analysis: evolution of research methods and topics, Artificial Intelligence Review (2023) 1–42.

[19] M. A. Alonso, D. Vilares, C. Gómez-Rodríguez, J. Vilares, Sentiment analysis for fake news detection, Electronics 10 (11) (2021) 1348.

[20] A. M. Luvembe, W. Li, S. Li, F. Liu, G. Xu, Dual emotion based fake news detection: A deep attention-weight update approach, Information Processing & Management 60 (4) (2023) 103354.

[21] C. Iwendi, S. Mohan, E. Ibeke, A. Ahmadian, T. Ciano, et al., Covid-19 fake news sentiment analysis, Computers and electrical engineering 101 (2022) 107967.

[22] C. Wang, B. Zhou, H. Tu, Y. Liu, Rumor detection on social media using temporal dynamic structure and emotional information, in: 2021 IEEE Sixth International Conference on Data Science in Cyberspace (DSC), IEEE, 2021, pp. 16–22.

[23] X. Miao, D. Rao, Z. Jiang, Syntax and sentiment enhanced bert for earliest rumor detection, in: Natural Language Processing and Chinese Computing: 10th CCF International Conference, NLPCC 2021, Qingdao, China, October 13–17, 2021, Proceedings, Part I 10, Springer, 2021, pp. 570–582.

[24] V. Kolev, G. Weiss, G. Spanakis, Foreal: Roberta model for fake news detection based on emotions., in: ICAART (2), 2022, pp. 429–440.

[25] B. Hu, Q. Sheng, J. Cao, Y. Shi, Y. Li, D. Wang, P. Qi, Bad actor, good advisor: Exploring the role of large language models in fake news detection, arXiv preprint arXiv:2309.12247 (2023).

[26] B. M. Pavlyshenko, Analysis of disinformation and fake news detection using fine-tuned large language model, arXiv preprint arXiv:2309.04704 (2023).

[27] T.-H. Cheung, K.-M. Lam, Factllama: Optimizing instruction-following language models with external knowledge for automated fact-checking, arXiv preprint arXiv:2309.00240 (2023).

[28] H. F. Al-Saif, H. Z. Al-Dossari, Exploring the role of emotions in arabic rumor detection in social media, Applied Sciences 13 (15) (2023) 8815.

[29] X. Zhang, J. Cao, X. Li, Q. Sheng, L. Zhong, K. Shu, Mining dual emotion for fake news detection, in: Proceedings of the web conference 2021, 2021, pp. 3465–3476.

Table 8
Emotion Detection Tools

Tool	Target Language	Description	Function
SenticNet [237]	English	Concept-level lexicon leveraging the denotative and connotative information associated with words and multi-word expressions.	sentiment, intensity, emotion
NRC Emotion Lexicon (EmoLex) [238]	English	Crowd-sourced lexicon associating words with emotions and sentiments.	sentiment, emotion
NRC Intensity Emotion Lexicon[194]	English	Lexicon associating words with real-valued intensity scores	emotion, intensity
AraNet [239]	Arabic	Collection of BERT-based social media processing tools predicting various types of information, including emotion, irony, and sentiment	sentiment, emotion, irony
CAMeL[240]	Arabic	Open-source package consisting of a set of Python APIs for NLP with accompanying command-line tools that thin-wrap these APIs	sentiment
Affective Lexicon Ontology (ALO)[198]	Chinese	Lexicon in which each entry is with an emotion and sentiment polarity	sentiment, emotion
TextBlob[a]	English	Python sentiment analysis library that uses the Natural Language ToolKit (NLTK)	sentiment scores with subject and polarity
LIWC[241]	Multilingual	Text analysis software to conduct various calculations related to emotions, social dynamics, and cognitive processes, by counting relevant words.	various text analyses
Valence Aware Dictionary for sEntiment Reasoning (VADER) [242]	English	Open-source rule-based sentiment analysis tool suitable for analyzing social media text	sentiment with score
Sentilex-PT02[b]	Portuguese	Sentiment lexicon for Portuguese, consisting of 7,014 lemmas and 82,347 inflected forms	sentiment
AFINN[243]	English	Open-source dictionary-based sentiment analysis tool, which assigns numerical sentiment polarity scores.	sentiment with score
cn-sentiment-measures[c]	Chinese	Toolkit for estimating Chinese sentiment scores based on multiple measures.	sentiment with score
EmoSenticNet (EmoSN)[244]	English	Enriched version of SenticNet, consisting of 13,189 words labeled according to Ekman's six basic emotions	sentiment, emotion
SentiStrength[63]	English	Sentiment strength detection algorithm that uses a lexical approach exploiting a list of sentiment-related terms	sentiment with strength
SentiWordNet[196]	English	Publicly available lexical resource that associates each WordNet synset with three numerical scores denoting objectivity, positivity, and negativity)	sentiment with score
HowNet[245]	Bilingual	Online common-sense knowledge base containing English and Chinese words that identifies inter-conceptual relations and inter-attribute relations of concepts	sentiment scores, sentimental words, degree words
SentiSense[246]	English	Lexicon that attaches 14 categories of emotional meanings to WordNet synsets.	sentiment, intensity, emotion
Affective Norms for English Words (ANEW) [247]	English	Lexicon of words rated by humans according to the Valence-Arousal-Dominance (VAD) model	Valence, Arousal, Dominance

[a] https://textblob.readthedocs.io/
[b] https://b2find9.cloud.dkrz.de/dataset/b6bd16c2-a8ab-598f-be41-1e7aeecd60d3
[c] https://github.com/dhchenx/cn-sentiment-measures

[30] A. Choudhry, I. Khatri, M. Jain, D. K. Vishwakarma, An emotion-aware multitask approach to fake news and rumor detection using transfer learning, IEEE Transactions on Computational Social Systems (2022) 1–12.

[31] L. Tian, X. Zhang, Y. Wang, H. Liu, Early detection of rumours on twitter via stance transfer learning, in: Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part I 42, Springer, 2020, pp. 575–588.

[32] C. Conforti, M. T. Pilehvar, N. Collier, Towards automatic fake news detection: cross-level stance detection in news articles, in: Proceedings of the first workshop on fact extraction and VERification (FEVER), 2018, pp. 40–49.

[33] L. Derczynski, K. Bontcheva, M. Liakata, R. Procter, G. W. S. Hoi, A. Zubiaga, Semeval-2017 task 8: Rumoureval: Determining rumour veracity and support for rumours, arXiv preprint arXiv:1704.05972 (2017).

[34] Z. Zojaji, B. Tork Ladani, Adaptive cost-sensitive stance classification model for rumor detection in social networks, Social Network Analysis and Mining 12 (1) (2022) 134.

[35] A. Upadhyaya, M. Fisichella, W. Nejdl, Towards sentiment and temporal aided stance detection of climate change tweets, Information Processing & Management 60 (4) (2023) 103325.

[36] A. Kim, P. L. Moravec, A. R. Dennis, Combating fake news on social media with source ratings: The effects of user and expert reputation ratings, Journal of Management Information Systems 36 (3) (2019) 931–968.

[37] A. Worrall, A. Ndumu, L. H. Gerido, Sentiment and network analysis of twitter reactions to the us birthright citizenship ban debate, in: International Conference on Information, Springer, 2022, pp. 149–174.

[38] S. M. Mohammad, P. Sobhani, S. Kiritchenko, Stance and sentiment in tweets, ACM Transactions on Internet Technology (TOIT) 17 (3) (2017) 1–23.

[39] P. Sobhani, S. Mohammad, S. Kiritchenko, Detecting stance in tweets and analyzing its interaction with sentiment, in: Proceedings of the fifth joint conference on lexical and computational semantics, 2016, pp. 159–169.

[40] A. Hanselowski, P. Avinesh, B. Schiller, F. Caspelherr, D. Chaudhuri, C. M. Meyer, I. Gurevych, A retrospective analysis of the fake news challenge stance detection task, in: Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 1859–1874.
[41] B. Ghanem, A. T. Cignarella, C. Bosco, P. Rosso, F. M. R. Pardo, Upv-28-unito at semeval-2019 task 7: Exploiting post's nesting and syntax information for rumor stance classification, in: Proceedings of the 13th international workshop on semantic evaluation, 2019, pp. 1125–1131.
[42] A. Khandelwal, Fine-tune longformer for jointly predicting rumor stance and veracity, in: Proceedings of the 3rd ACM India Joint International Conference on Data Science & Management of Data (8th ACM IKDD CODS & 26th COMAD), 2021, pp. 10–19.
[43] F. Wang, M. Lan, Y. Wu, Ecnu at semeval-2017 task 8: Rumour evaluation using effective features and supervised ensemble models, in: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), 2017, pp. 491–496.
[44] K. Xuan, R. Xia, Rumor stance classification via machine learning with text, user and propagation features, in: 2019 International Conference on Data Mining Workshops (ICDMW), IEEE, 2019, pp. 560–566.
[45] P. Parimi, R. R. Rout, Flacorm: fuzzy logic and ant colony optimization for rumor mitigation through stance prediction in online social networks, Social Network Analysis and Mining 13 (1) (2023) 22.
[46] Y. Liu, H. Shen, L. Shi, A review of rumor detection techniques in social networks, Journal of Intelligent & Fuzzy Systems (Preprint) (2023) 1–18.
[47] K. Mandal, A. Malik, A review on fake news detection techniques, in: 2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC), 2023, pp. 611–616. doi:10.1109/ICSCCC58608.2023.10176530.
[48] M. Tajrian, A. Rahman, M. A. Kabir, M. R. Islam, A review of methodologies for fake news analysis, IEEE Access 11 (2023) 73879–73893. doi:10.1109/ACCESS.2023.3294989.
[49] M. Choudhary, S. Jha, D. Saxena, A. K. Singh, et al., A review of fake news detection methods using machine learning, in: 2021 2nd International Conference for Emerging Technology (INCET), IEEE, 2021, pp. 1–5.
[50] A. Bondielli, F. Marcelloni, A survey on fake news and rumour detection techniques, Information Sciences 497 (2019) 38–55.
[51] I. Varlamis, D. Michail, F. Glykou, P. Tsantilas, A survey on the use of graph convolutional networks for combating fake news, Future Internet 14 (3) (2022) 70.
[52] A. D'Ulizia, M. C. Caschera, F. Ferri, P. Grifoni, Fake news detection: a survey of evaluation datasets, PeerJ Computer Science 7 (2021) e518.
[53] H. F. Alsaif, H. D. Aldossari, Review of stance detection for rumor verification in social media, Engineering Applications of Artificial Intelligence 119 (2023) 105801.
[54] M. Hardalov, A. Arora, P. Nakov, I. Augenstein, A survey on stance detection for mis-and disinformation identification, arXiv preprint arXiv:2103.00242 (2021).
[55] W. Shahid, Y. Li, D. Staples, G. Amin, S. Hakak, A. Ghorbani, Are you a cyborg, bot or human?—a survey on detecting fake news spreaders, IEEE Access 10 (2022) 27069–27083.
[56] S. Shelke, V. Attar, Source detection of rumor in social network–a review, Online Social Networks and Media 9 (2019) 30–42.
[57] T. Zhang, A. M. Schoene, S. Ji, S. Ananiadou, Natural language processing applied to mental illness detection: a narrative review, NPJ digital medicine 5 (1) (2022) 46.
[58] P. Przybyła, A. J. Brockmeier, G. Kontonatsios, M.-A. Le Pogam, J. McNaught, E. von Elm, K. Nolan, S. Ananiadou, Prioritising references for systematic reviews with robotanalyst: a user study, Research synthesis methods 9 (3) (2018) 470–488.
[59] A. O'Mara-Eves, J. Thomas, J. McNaught, M. Miwa, S. Ananiadou, Using text mining for study identification in systematic reviews: a systematic review of current approaches, Systematic reviews 4 (1) (2015) 1–22.
[60] M. Miwa, J. Thomas, A. O'Mara-Eves, S. Ananiadou, Reducing systematic review workload through certainty-based screening, Journal of biomedical informatics 51 (2014) 242–253.
[61] N. M. Hakak, M. Mohd, M. Kirmani, M. Mohd, Emotion analysis: A survey, in: 2017 International Conference on Computer, Communications and Electronics (Comptelix), 2017, pp. 397–402. doi:10.1109/COMPTELIX.2017.8004002.
[62] M. D. Devika, C. Sunitha, A. Ganesh, Sentiment analysis: a comparative study on different approaches, Procedia Computer Science 87 (2016) 44–49.
[63] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, A. Kappas, Sentiment strength detection in short informal text, Journal of the American society for information science and technology 61 (12) (2010) 2544–2558.
[64] H. A. Uymaz, S. K. Metin, Vector based sentiment and emotion analysis from text: A survey, Engineering Applications of Artificial Intelligence 113 (2022) 104922.
[65] K. Sailunaz, M. Dhaliwal, J. Rokne, R. Alhajj, Emotion detection from text and speech: a survey, Social Network Analysis and Mining 8 (2018) 1–26.
[66] P. Shaver, J. Schwartz, D. Kirson, C. O'connor, Emotion knowledge: further exploration of a prototype approach., Journal of personality and social psychology 52 (6) (1987) 1061.
[67] P. Ekman, An argument for basic emotions, Cognition & emotion 6 (3-4) (1992) 169–200.
[68] R. Plutchik, A general psychoevolutionary theory of emotion, in: Theories of emotion, Elsevier, 1980, pp. 3–33.
[69] F. A. Acheampong, C. Wenyu, H. Nunoo-Mensah, Text-based emotion detection: Advances, challenges, and opportunities, Engineering Reports 2 (7) (2020) e12189.
[70] J. A. Russell, A. Mehrabian, Evidence for a three-factor theory of emotions, Journal of research in Personality 11 (3) (1977) 273–294.
[71] A. B. Warriner, V. Kuperman, M. Brysbaert, Norms of valence, arousal, and dominance for 13,915 english lemmas, Behavior research methods 45 (2013) 1191–1207.
[72] A. Alslaity, R. Orji, Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions, Behaviour & Information Technology (2022) 1–26.
[73] S. Peng, L. Cao, Y. Zhou, Z. Ouyang, A. Yang, X. Li, W. Jia, S. Yu, A survey on deep learning for textual emotion analysis in social networks, Digital Communications and Networks 8 (5) (2022) 745–762.
[74] B. Zhang, H. Yang, T. Zhou, A. Babar, X.-Y. Liu, Enhancing financial sentiment analysis via retrieval augmented large language models, arXiv preprint arXiv:2310.04027 (2023).
[75] S. Feng, G. Sun, N. Lubis, C. Zhang, M. Gašić, Affect recognition in conversations using large language models, arXiv preprint arXiv:2309.12881 (2023).
[76] S. Lei, G. Dong, X. Wang, K. Wang, S. Wang, Instructerc: Reforming emotion recognition in conversation with a retrieval multi-task llms framework, arXiv preprint arXiv:2309.11911 (2023).
[77] L. Wu, F. Morstatter, X. Hu, H. Liu, Mining misinformation in social media, Big data in complex and social networks (2016) 123–152.
[78] R. K. Kaliyar, A. Goswami, P. Narang, Deepfake: improving fake news detection using tensor decomposition-based deep neural network, The Journal of Supercomputing 77 (2021) 1015–1037.
[79] V. Jain, R. K. Kaliyar, A. Goswami, P. Narang, Y. Sharma, Aenet: an attention-enabled neural architecture for fake news detection using contextual features, Neural Computing and Applications 34 (1) (2022) 771–782.
[80] R. K. Kaliyar, A. Goswami, P. Narang, S. Sinha, Fndnet–a deep convolutional neural network for fake news detection, Cognitive Systems Research 61 (2020) 32–44.
[81] Y. Wu, C. Zou, L. Wang, Z. Pan, Emotion makes rumor viral? the effects of discrete emotions on rumor-mongering on social media during a social crisis, WHICEB 2022 Proceedings (2022).

[82] C. Martel, G. Pennycook, D. G. Rand, Reliance on emotion promotes belief in fake news, Cognitive research: principles and implications 5 (2020) 1–20.

[83] A. Rijo, S. Waldzus, That's interesting! the role of epistemic emotions and perceived credibility in the relation between prior beliefs and susceptibility to fake-news, Computers in Human Behavior 141 (2023) 107619.

[84] W.-K. Tan, C. Y. Hsu, The application of emotions, sharing motivations, and psychological distance in examining the intention to share covid-19-related fake news, Online Information Review 47 (1) (2023) 59–80.

[85] K. Ali, C. Li, S. A. Muqtadir, et al., The effects of emotions, individual attitudes towards vaccination, and social endorsements on perceived fake news credibility and sharing motivations, Computers in human behavior 134 (2022) 107307.

[86] J. Li, L. Xiao, Multi-emotion recognition using multi-emobert and emotion analysis in fake news, in: Proceedings of the 15th ACM Web Science Conference 2023, 2023, pp. 128–135.

[87] M. Wan, Y. Zhong, X. Gao, S. Y. M. Lee, C.-R. Huang, Fake news, real emotions: Emotion analysis of covid-19 infodemic in weibo, IEEE Transactions on Affective Computing (2023).

[88] M. Prabhala, I. Bose, Do emotions determine rumors and impact the financial market? the case of demonetization in india, in: 2019 IEEE international conference on industrial engineering and engineering management (IEEM), IEEE, 2019, pp. 219–223.

[89] Y. Chuai, J. Zhao, Anger can make fake news viral online, Frontiers in Physics 10 (2022) 970174.

[90] L. Zhou, J. Tao, D. Zhang, Does fake news in different languages tell the same story? an analysis of multi-level thematic and emotional characteristics of news about covid-19, Information Systems Frontiers 25 (2) (2023) 493–512.

[91] L. Zhou, J. Tao, D. Zhang, Does fake news in different languages tell the same story? an analysis of multi-level thematic and emotional characteristics of news about covid-19, Information Systems Frontiers 25 (2) (2023) 493–512.

[92] K. Solovev, N. Pröllochs, Moral emotions shape the virality of covid-19 misinformation on social media, in: Proceedings of the ACM web conference 2022, 2022, pp. 3706–3717.

[93] S. Goel, A. Anderson, J. Hofman, D. J. Watts, The structural virality of online diffusion, Management Science 62 (1) (2016) 180–196.

[94] M. Sui, I. Hawkins, R. Wang, When falsehood wins? varied effects of sensational elements on users' engagement with real and fake posts, Computers in Human Behavior 142 (2023) 107654.

[95] E. Fersini, J. Armanini, M. D'Intorni, et al., Profiling fake news spreaders: Stylometry, personality, emotions and embeddings., in: CLEF (Working Notes), 2020.

[96] . CLEF2020, Clef2020 - checkthat! lab, in: CLEF, 2020. URL https://sites.google.com/view/clef2020-checkthat/

[97] P. Wang, H. Shi, X. Wu, L. Jiao, Sentiment analysis of rumor spread amid covid-19: Based on weibo text, in: Healthcare, Vol. 9, MDPI, 2021, p. 1275.

[98] N. S. Khan, F. Molla, R. S. Khan, E. H. Shamim, S. Hossain, M. M. Hasan, Exploration of online fake news through machine learning and sentiment analyses, in: International Conference on Intelligent Computing & Optimization, Springer, 2022, pp. 439–448.

[99] M. Gagiano, V. Marivate, Emotionally driven fake news in south africa, EPiC Series in Computing 93 (2023) 56–67.

[100] P. Patwa, S. Sharma, S. Pykl, V. Guptha, G. Kumari, M. S. Akhtar, A. Ekbal, A. Das, T. Chakraborty, Fighting an infodemic: Covid-19 fake news dataset, in: Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1, Springer, 2021, pp. 21–29.

[101] L. Cui, D. Lee, Coaid: Covid-19 healthcare misinformation dataset, arXiv preprint arXiv:2006.00885 (2020).

[102] S. Li, Explore covid-19 infodemic, in: Towards Data Science, 2020. URL https://towardsdatascience.com/explore-covid-19-infodemic-2d1ceaae2306

[103] S. Arora, R. Rani, N. Saxena, Modified valence aware dictionary for sentiment reasoning classifier for detection and classification of covid-19 related rumors from social media data streams, Concurrency and Computation: Practice and Experience 34 (21) (2022) e7124.

[104] E. Kochkina, M. Liakata, A. Zubiaga, PHEME dataset for Rumour Detection and Veracity Classification (6 2018). doi:10.6084/m9.figshare.6392078.v1. URL https://figshare.com/articles/dataset/PHEME_dataset_for_Rumour_Detection_and_Veracity_Classification/6392078

[105] J. Ma, W. Gao, K.-F. Wong, Detect rumors in microblog posts using propagation structure via kernel learning, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vol. 1, 2017, pp. 708–717.

[106] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, C. Meeyoung, Detecting rumors from microblogs with recurrent neural networks, in: The 25th International Joint Conference on Artificial Intelligence, AAAI, 2016.

[107] Q. Nan, J. Cao, Y. Zhu, Y. Wang, J. Li, Mdfend: Multi-domain fake news detection, in: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, 2021, pp. 3343–3347.

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

[109] K. Nakamura, S. Levy, W. Y. Wang, Fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection, in: Proceedings of the Twelfth Language Resources and Evaluation Conference, 2020, pp. 6149–6157.

[110] C. Boididou, S. Papadopoulos, D. T. Dang Nguyen, G. Boato, M. Riegler, A. Petlund, I. Kompatsiaris, Verifying multimedia use at mediaeval 2016, 2016.

[111] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, R. Mihalcea, Automatic detection of fake news, in: Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 3391–3401.

[112] H. Ahmed, I. Traore, S. Saad, Detecting opinion spams and fake news using text classification, Security and Privacy 1 (1) (2018) e9.

[113] H. Ahmed, I. Traore, S. Saad, 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 1, Springer, 2017, pp. 127–138.

[114] W. Y. Wang, "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 (Volume 2: Short Papers), Association for Computational Linguistics, Vancouver, Canada, 2017, pp. 422–426. URL https://aclanthology.org/P17-2067

[115] T. Alhindi, S. Petridis, S. Muresan, Where is your evidence: Improving fact-checking by justification modeling, in: Proceedings of the first workshop on fact extraction and verification (FEVER), 2018, pp. 85–90.

[116] T. Mitra, E. Gilbert, Credbank: A large-scale social media corpus with associated credibility annotations, in: Proceedings of the international AAAI conference on web and social media, Vol. 9, 2015, pp. 258–267.

[117] . Kaggle, Getting real about fake news, in: kaggle, 2016. URL https://www.kaggle.com/datasets/mrisdal/fake-news

[118] V. L. Rubin, N. Conroy, Y. Chen, S. Cornwell, Fake news or truth? using satirical cues to detect potentially misleading news, in: Proceedings of the second workshop on computational approaches to deception detection, 2016, pp. 7–17.

[119] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, Y. Choi, Truth of varying shades: Analyzing language in fake news and political fact-checking, in: Proceedings of the 2017 conference on empirical methods in

natural language processing, 2017, pp. 2931–2937.
[120] S. Kwon, M. Cha, K. Jung, Rumor detection over varying time windows, PloS one 12 (1) (2017) e0168344.
[121] R. Sicilia, S. L. Giudice, Y. Pei, M. Pechenizkiy, P. Soda, Twitter rumour detection in the health domain, Expert Systems with Applications 110 (2018) 33–40.
[122] B. Ghanem, S. P. Ponzetto, P. Rosso, F. Rangel, Fakeflow: Fake news detection by modeling the flow of affective information, in: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, 2021, pp. 679–689.
[123] S. Castelo, T. Almeida, A. Elghafari, A. Santos, K. Pham, E. Nakamura, J. Freire, A topic-agnostic approach for identifying fake news pages, in: Companion proceedings of the 2019 World Wide Web conference, 2019, pp. 975–980.
[124] B. Horne, S. Adali, This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news, in: Proceedings of the international AAAI conference on web and social media, Vol. 11, 2017, pp. 759–766.
[125] S. Rezaei, M. Kahani, B. Behkamal, A. Jalayer, Early multi-class ensemble-based fake news detection using content features, Social Network Analysis and Mining 13 (1) (2022) 16.
[126] K. Anoop, P. Deepak, V. Lajish, Emotion cognizance improves health fake news identification., in: IDEAS, Vol. 2020, 2020, p. 24th.
[127] W. Cuenca, C. González-Fernández, A. Fernández-Isabel, I. Martín de Diego, A. G. Martín, Combining conceptual graphs and sentiment analysis for fake news detection, in: Computational Intelligence and Mathematics for Tackling Complex Problems 2, Springer, 2022, pp. 129–138.
[128] I. Augenstein, C. Lioma, D. Wang, L. C. Lima, C. Hansen, C. Hansen, J. G. Simonsen, Multifc: A real-world multi-domain dataset for evidence-based fact checking of claims, arXiv preprint arXiv:1909.03242 (2019).
[129] Z. Yang, J. Ma, H. Chen, H. Lin, Z. Luo, Y. Chang, A coarse-to-fine cascaded evidence-distillation neural network for explainable fake news detection, arXiv preprint arXiv:2209.14642 (2022).
[130] M. Chen, L. Wei, H. Cao, W. Zhou, S. Hu, Can large language models understand content and propagation for misinformation detection: An empirical study, arXiv preprint arXiv:2311.12699 (2023).
[131] Z. Ma, M. Liu, G. Fang, Y. Shen, Ltcr: Long-text chinese rumor detection dataset, arXiv preprint arXiv:2306.07201 (2023).
[132] R. Kumari, N. Ashok, P. K. Agrawal, T. Ghosal, A. Ekbal, Identifying multimodal misinformation leveraging novelty detection and emotion recognition, Journal of Intelligent Information Systems (2023) 1–22.
[133] V. Gupta, R. Kumari, N. Ashok, T. Ghosal, A. Ekbal, Mmm: An emotion and novelty-aware approach for multilingual multimodal misinformation detection, in: Findings of the Association for Computational Linguistics: AACL-IJCNLP 2022, 2022, pp. 464–477.
[134] Z. Jin, J. Cao, H. Guo, Y. Zhang, J. Luo, Multimodal fusion with recurrent neural networks for rumor detection on microblogs, in: Proceedings of the 25th ACM international conference on Multimedia, 2017, pp. 795–816.
[135] C. Guo, J. Cao, X. Zhang, K. Shu, M. Yu, Exploiting emotions for fake news detection on social media, arXiv preprint arXiv:1903.01728 (2019).
[136] F. C. D. da Silva, Sentiment gradient, an enhancement to the truth, lies and sarcasm detection, in: Advances in Artificial Intelligence–IBERAMIA 2022: 17th Ibero-American Conference on AI, Cartagena de Indias, Colombia, November 23–25, 2022, Proceedings, Vol. 13788, Springer Nature, 2023, p. 107.
[137] F. R. M. da Silva, P. M. S. Freire, M. P. de Souza, G. de AB Plenamente, R. R. Goldschmidt, Fakenewssetgen: A process to build datasets that support comparison among fake news detection methods, in: Proceedings of the Brazilian Symposium on Multimedia and the Web, 2020, pp. 241–248.
[138] R. A. Monteiro, R. L. Santos, T. A. Pardo, T. A. De Almeida, E. E. Ruiz, O. A. Vale, Contributions to the study of fake news in portuguese: New corpus and automatic detection results, in: Computational Processing of the Portuguese Language: 13th International Conference, PROPOR 2018, Canela, Brazil, September 24–26, 2018, Proceedings 13, Springer, 2018, pp. 324–334.
[139] A. Barrón-Cedeno, T. Elsayed, P. Nakov, G. Da San Martino, M. Hasanain, R. Suwaileh, F. Haouari, N. Babulkov, B. Hamdan, A. Nikolov, et al., Overview of checkthat! 2020: Automatic identification and verification of claims in social media, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction: 11th International Conference of the CLEF Association, CLEF 2020, Thessaloniki, Greece, September 22–25, 2020, Proceedings 11, Springer, 2020, pp. 215–236.
[140] A. E. Lillie, E. R. Middelboe, L. Derczynski, Joint rumour stance and veracity prediction, in: Nordic Conference of Computational Linguistics (2019), Linköping University Electronic Press, 2019, pp. 208–221.
[141] . Bytedance WSDM Cup, J. Qi, Wsdm - fake news classification, in: Kaggle, 2018.
URL https://kaggle.com/competitions/fake-news-pair-classification-challenge
[142] G. Gorrell, E. Kochkina, M. Liakata, A. Aker, A. Zubiaga, K. Bontcheva, L. Derczynski, Semeval-2019 task 7: Rumoureval 2019: Determining rumour veracity and support for rumours, in: Proceedings of the 13th International Workshop on Semantic Evaluation: NAACL HLT 2019, Association for Computational Linguistics, 2019, pp. 845–854.
[143] W. Ferreira, A. Vlachos, Emergent: a novel data-set for stance classification, in: Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies, ACL, 2016.
[144] . Delip Rao Dean Pomerleau, J. Qi, Fake news challenge, 2017.
URL http://www.fakenewschallenge.org/
[145] E. C. Mutlu, T. Oghaz, J. Jasser, E. Tutunculer, A. Rajabi, A. Tayebi, O. Ozmen, I. Garibay, A stance data set on polarized conversations on twitter about the efficacy of hydroxychloroquine as a treatment for covid-19, Data in brief 33 (2020) 106401.
[146] Z. Arkaitz, L. Maria, P. Rob, W. S. H. Geraldine, T. Peter, Pheme rumour scheme dataset: journalism use case, 2016.
URL https://api.semanticscholar.org/CorpusID:130779128
[147] M. Lukasik, T. Cohn, K. Bontcheva, Classifying tweet level judgements of rumours in social media, arXiv preprint arXiv:1506.00468 (2015).
[148] G. Giasemidis, C. Singleton, I. Agrafiotis, J. R. Nurse, A. Pilgrim, C. Willis, D. V. Greetham, Determining the veracity of rumours on twitter, in: Social Informatics: 8th International Conference, SocInfo 2016, Bellevue, WA, USA, November 11-14, 2016, Proceedings, Part I 8, Springer, 2016, pp. 185–205.
[149] K. Kawintiranon, L. Singh, Knowledge enhanced masked language model for stance detection, in: Proceedings of the 2021 conference of the north american chapter of the association for computational linguistics: human language technologies, 2021, pp. 4725–4735.
[150] L. Zeng, K. Starbird, E. Spiro, # unconfirmed: Classifying rumor stance in crisis-related social media messages, in: Proceedings of the international aaai conference on web and social media, Vol. 10, 2016, pp. 747–750.
[151] L. Cui, S. Wang, D. Lee, Same: sentiment-aware multi-modal embedding for detecting fake news, in: Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining, 2019, pp. 41–48.
[152] B. Bhutani, N. Rastogi, P. Sehgal, A. Purwar, Fake news detection using sentiment analysis, in: 2019 twelfth international conference on contemporary computing (IC3), IEEE, 2019, pp. 1–5.
[153] O. Ajao, D. Bhowmik, S. Zargari, Sentiment aware fake news detection on online social networks, in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 2507–2511.
[154] Z. Wang, Y. Guo, J. Wang, Z. Li, M. Tang, Rumor events detection from chinese microblogs via sentiments enhancement, IEEE Access

7 (2019) 103000–103018.

[155] Z. Wang, Y. Guo, Rumor events detection enhanced by encoding sentimental information into time series division and word representations, Neurocomputing 397 (2020) 224–243.

[156] S. Dong, Z. Qian, P. Li, X. Zhu, Q. Zhu, Rumor detection on hierarchical attention network with user and sentiment information, in: Natural Language Processing and Chinese Computing: 9th CCF International Conference, NLPCC 2020, Zhengzhou, China, October 14–18, 2020, Proceedings, Part II 9, Springer, 2020, pp. 366–377.

[157] M. P. de Souza, F. R. M. da Silva, P. M. S. Freire, R. R. Goldschmidt, A linguistic-based method that combines polarity, emotion and grammatical characteristics to detect fake news in portuguese, in: Proceedings of the Brazilian Symposium on Multimedia and the Web, 2020, pp. 217–224.

[158] L. Ding, L. Ding, R. O. Sinnott, Fake news classification of social media through sentiment analysis, in: Big Data–BigData 2020: 9th International Conference, Held as Part of the Services Conference Federation, SCF 2020, Honolulu, HI, USA, September 18-20, 2020, Proceedings 9, Springer, 2020, pp. 52–67.

[159] I. Touahri, A. Mazroui, Evolutionteam at clef2020-checkthat! lab: Integration of linguistic and sentimental features in a fake news detection approach., in: CLEF (Working Notes), 2020.

[160] U. Ezeakunne, S. M. Ho, X. Liu, Sentiment and retweet analysis of user response for early fake news detection, in: The International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation (SBP-BRiMS'20), 2020, pp. 1–10.

[161] R. Kumari, N. Ashok, T. Ghosal, A. Ekbal, A multitask learning approach for fake news detection: Novelty, emotion, and sentiment lend a helping hand, in: 2021 International Joint Conference on Neural Networks (IJCNN), IEEE, 2021, pp. 1–8.

[162] I. M. L. Maia, M. P. de Souza, F. R. M. da Silva, P. M. S. Freire, R. R. Goldschmidt, A sentiment-based multimodal method to detect fake news, in: Proceedings of the Brazilian Symposium on Multimedia and the Web, 2021, pp. 213–216.

[163] D. Dong, F. Lin, G. Li, B. Liu, Sentiment-aware fake news detection on social media with hypergraph attention networks, in: 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2022, pp. 2174–2180.

[164] I. Kelk, B. Basseri, W. Lee, R. Qiu, C. Tanner, Automatic fake news detection: Are current models "fact-checking" or "gut-checking"?, in: Proceedings of the Fifth Fact Extraction and VERification Workshop (FEVER), 2022, pp. 29–36.

[165] B. Mohamed, H. Haytam, F. Abdelhadi, Applying fuzzy logic and neural network in sentiment analysis for fake news detection: case of covid-19, Combating fake news with computational intelligence techniques (2022) 387–400.

[166] A. Haque, M. Abulaish, A graph-based approach leveraging posts and reactions for detecting rumors on online social media, in: Proceedings of the 36th Pacific Asia Conference on Language, Information and Computation, 2022, pp. 533–544.

[167] S. K. Uppada, K. Manasa, B. Vidhathri, R. Harini, B. Sivaselvan, Novel approaches to fake news and fake account detection in osns: user social engagement and visual content centric model, Social Network Analysis and Mining 12 (1) (2022) 52.

[168] S. K. Uppada, P. Patel, An image and text-based multimodal model for detecting fake news in osn's, Journal of Intelligent Information Systems (2022) 1–27.

[169] R. Kumari, N. Ashok, T. Ghosal, A. Ekbal, What the fake? probing misinformation detection standing on the shoulder of novelty and emotion, Information Processing & Management 59 (1) (2022) 102740.

[170] C. Fu, K. Chen, X. Pan, S. Yu, J. Ni, Y. Min, Rumor detection based on the temporal sentiment, in: China National Conference on Big Data and Social Computing, Springer, 2022, pp. 275–290.

[171] N. Seddari, A. Derhab, M. Belaoued, W. Halboob, J. Al-Muhtadi, A. Bouras, A hybrid linguistic and knowledge-based analysis approach for fake news detection on social media, IEEE Access 10 (2022) 62097–62109.

[172] S. V. Balshetwar, A. Rs, Fake news detection in social media based on sentiment analysis using classifier techniques, Multimedia Tools and Applications (2023) 1–31.

[173] S. K. Hamed, M. J. Ab Aziz, M. R. Yaakub, Fake news detection model on social media by leveraging sentiment analysis of news content and emotion analysis of users' comments, Sensors 23 (4) (2023) 1748.

[174] Q. Guo, Z. Kang, L. Tian, Z. Chen, Tiefake: Title-text similarity and emotion-aware fake news detection, arXiv preprint arXiv:2304.09421 (2023).

[175] L. Fang, K. Feng, K. Zhao, A. Hu, T. Li, Unsupervised rumor detection based on propagation tree vae, IEEE Transactions on Knowledge and Data Engineering (2023).

[176] H. Zhang, Z. Li, S. Liu, T. Huang, Z. Ni, J. Zhang, Z. Lv, Do sentence-level sentiment interactions matter? sentiment mixed heterogeneous network for fake news detection, IEEE Transactions on Computational Social Systems (2023).

[177] G. Ali, M. S. I. Malik, Rumour identification on twitter as a function of novel textual and language-context features, Multimedia Tools and Applications 82 (5) (2023) 7017–7038.

[178] J. Zhao, Z. Zhao, L. Shi, Z. Kuang, Y. Liu, Collaborative mixture-of-experts model for multi-domain fake news detection, Electronics 12 (16) (2023) 3440.

[179] X. Zhang, Y. Pan, X. Gu, G. Liang, Sentiment analysis-based social network rumor detection model with bi-directional graph convolutional networks, in: International Conference on Computer Application and Information Security (ICCAIS 2022), Vol. 12609, SPIE, 2023, pp. 463–469.

[180] S. E. V. S. Pillai, W.-C. Hu, Misinformation detection using an ensemble method with emphasis on sentiment and emotional analyses, in: 2023 IEEE/ACIS 21st International Conference on Software Engineering Research, Management and Applications (SERA), IEEE, 2023, pp. 295–300.

[181] F. Barbieri, J. Camacho-Collados, L. Neves, L. Espinosa-Anke, Tweeteval: Unified benchmark and comparative evaluation for tweet classification, arXiv preprint arXiv:2010.12421 (2020).

[182] A. Choudhry, I. Khatri, A. Chakraborty, D. Vishwakarma, M. Prasad, Emotion-guided cross-domain fake news detection using adversarial domain adaptation, in: Proceedings of the 19th International Conference on Natural Language Processing (ICON), 2022, pp. 75–79.

[183] A. Chakraborty, I. Khatri, A. Choudhry, P. Gupta, D. K. Vishwakarma, M. Prasad, An emotion-guided approach to domain adaptive fake news detection using adversarial learning (student abstract), in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 37, 2023, pp. 16178–16179.

[184] K. Sharifani, M. Amini, Machine learning and deep learning: A review of methods and applications, World Information Technology and Engineering Journal 10 (07) (2023) 3897–3904.

[185] M. H. Al-Tai, B. M. Nema, A. Al-Sherbaz, Deep learning for fake news detection: Literature review, Al-Mustansiriyah Journal of Science 34 (2) (2023) 70–81.

[186] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, E. Hovy, Hierarchical attention networks for document classification, in: Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, 2016, pp. 1480–1489.

[187] S. Sabour, N. Frosst, G. E. Hinton, Dynamic routing between capsules, Advances in neural information processing systems 30 (2017).

[188] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, S. Ravi, Goemotions: A dataset of fine-grained emotions, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 4040–4054.

[189] Y. Li, H. Su, X. Shen, W. Li, Z. Cao, S. Niu, Dailydialog: A manually labelled multi-turn dialogue dataset, arXiv preprint arXiv:1710.03957 (2017).

[190] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[191] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).

[192] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1251–1258.

[193] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al., Learning transferable visual models from natural language supervision, in: International conference on machine learning, PMLR, 2021, pp. 8748–8763.

[194] S. M. Mohammad, Word affect intensities, in: Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018), Miyazaki, Japan, 2018.

[195] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781 (2013).

[196] A. Esuli, F. Sebastiani, Sentiwordnet: a high-coverage lexical resource for opinion mining, Evaluation 17 (1) (2007) 26.

[197] S. Zhang, L. Wang, K. Sun, X. Xiao, A practical chinese dependency parser based on a large-scale dataset, arXiv preprint arXiv:2009.00901 (2020).

[198] L. Xu, H. Lin, Y. Pan, H. Ren, J. Chen, Constructing the affective lexicon ontology, Journal of the China society for scientific and technical information 27 (2) (2008) 180–185.

[199] L. Zadeh, M. Gupta, R. Ragade, R. Yager, Advances in fuzzy set theory and applications, Gupta, M (1979) 318.

[200] N. Colnerič, J. Demšar, Emotion recognition on twitter: Comparative study and training a unison model, IEEE transactions on affective computing 11 (3) (2018) 433–446.

[201] L. H. Li, M. Yatskar, D. Yin, C.-J. Hsieh, K.-W. Chang, Visualbert: A simple and performant baseline for vision and language, arXiv preprint arXiv:1908.03557 (2019).

[202] H. Bahuleyan, O. Vechtomova, Uwaterloo at semeval-2017 task 8: Detecting stance towards rumours with topic independent features, in: Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017), 2017, pp. 461–464.

[203] A. Aker, L. Derczynski, K. Bontcheva, Simple open stance classification for rumour analysis, in: Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017, INCOMA Ltd., Varna, Bulgaria, 2017, pp. 31–39. doi:10.26615/978-954-452-049-6_005.
URL https://doi.org/10.26615/978-954-452-049-6_005

[204] O. Enayet, S. R. El-Beltagy, Niletmrg at semeval-2017 task 8: Determining rumour and veracity support for rumours on twitter., in: Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017), 2017, pp. 470–474.

[205] A. Srivastava, G. Rehm, J. M. Schneider, Dfki-dkt at semeval-2017 task 8: Rumour detection and classification using cascading heuristics, in: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), 2017, pp. 486–490.

[206] R. Masood, A. Aker, The fake news challenge: Stance detection using traditional machine learning approaches., in: KMIS, 2018, pp. 126–133.

[207] G. Bhatt, A. Sharma, S. Sharma, A. Nagpal, B. Raman, A. Mittal, Combining neural, statistical and external features for fake news stance identification, in: Companion Proceedings of the The Web Conference 2018, 2018, pp. 1353–1357.

[208] A. Janchevski, S. Gievska, Andrejjan at semeval-2019 task 7: A fusion approach for exploring the key factors pertaining to rumour analysis, in: Proceedings of the 13th International Workshop on Semantic Evaluation, 2019, pp. 1083–1089.

[209] S. Hamidian, M. Diab, Gwu nlp at semeval-2019 task 7: Hybrid pipeline for rumour veracity and stance classification on social media, in: Proceedings of the 13th international workshop on semantic evaluation, 2019, pp. 1115–1119.

[210] E. W. Pamungkas, V. Basile, V. Patti, Stance classification for rumour analysis in twitter: Exploiting affective information and conversation structure, arXiv preprint arXiv:1901.01911 (2019).

[211] G. Giasemidis, N. Kaplis, I. Agrafiotis, J. R. Nurce, A semi-supervised approach to message stance classification, IEEE Transactions on Knowledge and Data Engineering 32 (1) (2020) 1–11.

[212] C. Baziotis, N. Pelekis, C. Doulkeridis, Datastories at semeval-2017 task 4: Deep lstm with attention for message-level and topic-based sentiment analysis, in: Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017), 2017, pp. 747–754.

[213] L. Wu, Y. Rao, Y. Zhao, H. Liang, A. Nazir, Dtca: Decision tree-based co-attention networks for explainable claim verification, arXiv preprint arXiv:2004.13455 (2020).

[214] R. K. Kaliyar, P. Kumar, M. Kumar, M. Narkhede, S. Namboodiri, S. Mishra, Deepnet: an efficient neural network for fake news detection using news-user engagements, in: 2020 5th International Conference on Computing, Communication and Security (ICCCS), IEEE, 2020, pp. 1–6.

[215] C. Ma, X. Zhang, Gf-vae: a flow-based variational autoencoder for molecule generation, in: Proceedings of the 30th ACM international conference on information & knowledge management, 2021, pp. 1181–1190.

[216] C.-H. Shih, B.-C. Yan, S.-H. Liu, B. Chen, Investigating siamese lstm networks for text categorization, in: 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), IEEE, 2017, pp. 641–646.

[217] K. M. Caramancion, News verifiers showdown: A comparative performance evaluation of chatgpt 3.5, chatgpt 4.0, bing ai, and bard in news fact-checking, arXiv preprint arXiv:2306.17176 (2023).

[218] C. Yang, P. Zhang, W. Qiao, H. Gao, J. Zhao, Rumor detection on social media with crowd intelligence and chatgpt-assisted networks, in: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 2023, pp. 5705–5717.

[219] Q. Li, Q. Zhang, L. Si, eventai at semeval-2019 task 7: Rumor detection on social media by exploiting content, user credibility and propagation information, in: Proceedings of the 13th international workshop on semantic evaluation, 2019, pp. 855–859.

[220] R. Mao, Q. Liu, K. He, W. Li, E. Cambria, The biases of pre-trained language models: An empirical study on prompt-based sentiment analysis and emotion detection, IEEE Transactions on Affective Computing (2022).

[221] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, et al., Llama: Open and efficient foundation language models, arXiv preprint arXiv:2302.13971 (2023).

[222] C. Comito, L. Caroprese, E. Zumpano, Multimodal fake news detection on social media: a survey of deep learning techniques, Social Network Analysis and Mining 13 (1) (2023) 1–22.

[223] H. Liu, C. Li, Q. Wu, Y. J. Lee, Visual instruction tuning, arXiv preprint arXiv:2304.08485 (2023).

[224] H. Liu, C. Li, Y. Li, Y. J. Lee, Improved baselines with visual instruction tuning, arXiv preprint arXiv:2310.03744 (2023).

[225] K. Zheng, X. He, X. E. Wang, Minigpt-5: Interleaved vision-and-language generation via generative vokens, arXiv preprint arXiv:2310.02239 (2023).

[226] P. Qi, Y. Bu, J. Cao, W. Ji, R. Shui, J. Xiao, D. Wang, T.-S. Chua, Fakesv: A multimodal benchmark with rich social context for fake news detection on short video platforms, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 37, 2023, pp. 14444–14452.

[227] M. Hosseini, A. J. Sabet, S. He, D. Aguiar, Interpretable fake news detection with topic and deep variational models, Online Social Networks and Media 36 (2023) 100249.

[228] V. Dua, A. Rajpal, S. Rajpal, M. Agarwal, N. Kumar, I-flash: Interpretable fake news detector using lime and shap, Wireless Personal Communications (2023) 1–34.
[229] X. Gao, W. Chen, L. Lu, Y. Cui, X. Dai, L. Dai, K. Wang, J. Shen, Y. Wang, S. Wang, et al., An interpretable fake news detection method based on commonsense knowledge graph, Applied Sciences 13 (11) (2023) 6680.
[230] H. Zhao, H. Chen, F. Yang, N. Liu, H. Deng, H. Cai, S. Wang, D. Yin, M. Du, Explainability for large language models: A survey, arXiv preprint arXiv:2309.01029 (2023).
[231] K. Yang, T. Zhang, Z. Kuang, Q. Xie, S. Ananiadou, Mentalllama: Interpretable mental health analysis on social media with large language models, arXiv preprint arXiv:2309.13567 (2023).
[232] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, et al., A survey of large language models, arXiv preprint arXiv:2303.18223 (2023).
[233] L. He, T. Yin, K. Zheng, They may not work! an evaluation of eleven sentiment analysis tools on seven social media datasets, Journal of Biomedical Informatics 132 (2022) 104142.
[234] J. Pennington, R. Socher, C. D. Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543.
[235] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).
[236] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, Advances in neural information processing systems 26 (2013).
[237] E. Cambria, Q. Liu, S. Decherchi, F. Xing, K. Kwok, Senticnet 7: A commonsense-based neurosymbolic ai framework for explainable sentiment analysis, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference, 2022, pp. 3829–3839.
[238] S. M. Mohammad, P. D. Turney, Crowdsourcing a word–emotion association lexicon, Computational intelligence 29 (3) (2013) 436–465.
[239] M. Abdul-Mageed, C. Zhang, A. Hashemi, E. M. B. Nagoudi, Aranet: A deep learning toolkit for arabic social media, arXiv preprint arXiv:1912.13072 (2019).
[240] O. Obeid, N. Zalmout, S. Khalifa, D. Taji, M. Oudah, B. Alhafni, G. Inoue, F. Eryani, A. Erdmann, N. Habash, Camel tools: An open source python toolkit for arabic natural language processing, in: Proceedings of the Twelfth Language Resources and Evaluation Conference, 2020, pp. 7022–7032.
[241] J. W. Pennebaker, R. L. Boyd, K. Jordan, K. Blackburn, The development and psychometric properties of liwc2015, Tech. rep. (2015).
[242] C. Hutto, E. Gilbert, Vader: A parsimonious rule-based model for sentiment analysis of social media text, in: Proceedings of the international AAAI conference on web and social media, Vol. 8, 2014, pp. 216–225.
[243] F. Å. Nielsen, A new anew: Evaluation of a word list for sentiment analysis in microblogs, arXiv preprint arXiv:1103.2903 (2011).
[244] S. Poria, A. Gelbukh, A. Hussain, N. Howard, D. Das, S. Bandyopadhyay, Enhanced senticnet with affective labels for concept-based opinion mining, IEEE Intelligent Systems 28 (2) (2013) 31–38.
[245] Z. Dong, Q. Dong, Hownet-a hybrid language and knowledge resource, in: International conference on natural language processing and knowledge engineering, 2003. Proceedings. 2003, IEEE, 2003, pp. 820–824.
[246] J. C. De Albornoz, L. Plaza, P. Gervás, Sentisense: An easily scalable concept-based affective lexicon for sentiment analysis., in: LREC, Vol. 12, 2012, pp. 3562–3567.
[247] M. M. Bradley, P. J. Lang, Affective norms for english words (anew): Instruction manual and affective ratings, Tech. rep., Technical report C-1, the center for research in psychophysiology … (1999).
[248] F. Carvalho, G. Santos, G. P. Guedes, Affectpt-br: an affective lexicon based on liwc 2015, in: 2018 37th International Conference of the Chilean Computer Science Society (SCCC), IEEE, 2018, pp. 1–5.
[249] B. Liu, M. Hu, J. Cheng, Opinion observer: analyzing and comparing opinions on the web, in: Proceedings of the 14th international conference on World Wide Web, 2005, pp. 342–351.
[250] C. Hansen, C. Hansen, S. Alstrup, J. Grue Simonsen, C. Lioma, Neural check-worthiness ranking with weak supervision: Finding sentences for fact-checking, in: Companion proceedings of the 2019 world wide web conference, 2019, pp. 994–1000.
[251] E. Fast, B. Chen, M. S. Bernstein, Empath: Understanding topic signals in large-scale text, in: Proceedings of the 2016 CHI conference on human factors in computing systems, 2016, pp. 4647–4657.
[252] Y. Choi, J. Wiebe, +/-effectwordnet: Sense-level lexicon acquisition for opinion inference, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1181–1191.
[253] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, B. Qin, Learning sentiment-specific word embedding for twitter sentiment classification, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2014, pp. 1555–1565.
[254] S. M. Mohammad, S. Kiritchenko, X. Zhu, Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets, arXiv preprint arXiv:1308.6242 (2013).
[255] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, C. Potts, Recursive deep models for semantic compositionality over a sentiment treebank, in: Proceedings of the 2013 conference on empirical methods in natural language processing, 2013, pp. 1631–1642.
[256] C. Whissell, Using the revised dictionary of affect in language to quantify the emotional undertones of samples of natural language, Psychological reports 105 (2) (2009) 509–521.

1. A survey of emotion-based misinformation detection is provided.
2. Different advanced fusion methods are summarized.
3. The available datasets are categorized and summarized.
4. The performance of fusion methods is discussed.
5. The challenges and future directions for detecting misinformation are presented.

Declaration of interests

☒ 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 paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: