Social Network Analysis and Mining (2024) 14:176
https://doi.org/10.1007/s13278-024-01344-4

REVIEW

Fake news detection: recent trends and challenges

Hemang Thakar[1,2] · Brijesh Bhatt[2]

Received: 9 December 2023 / Revised: 16 July 2024 / Accepted: 20 August 2024

Abstract
The proliferation of fake news in the digital age has spurred extensive research efforts toward developing effective detection techniques. This abstract delves into recent trends and challenges within the domain of fake news detection. The ubiquity of social media platforms and user-generated content has led to the rapid dissemination of misinformation, necessitating robust mechanisms for differentiating between authentic and fabricated news. This paper explores emerging approaches, such as advanced machine learning models, natural language processing techniques, and cross-modal analysis, which leverage textual, visual, and contextual cues to enhance detection accuracy. However, as fake news tactics become more sophisticated, challenges like adversarial attacks, data scarcity, and domain adaptation come to the forefront. This abstract highlights the ongoing efforts to address these challenges and emphasizes the importance of interdisciplinary collaboration to devise comprehensive solutions for combating the intricate landscape of fake news dissemination.

Keywords Fake news · Social media · Data · Online context · Machine learning

1 Introduction

In an era characterized by the rapid dissemination of information through digital platforms, the proliferation of fake news has emerged as a critical concern. The term”fake news” refers to intentionally fabricated or misleading information presented as genuine news, often designed to deceive (Jain et al. 2022), manipulate, or exploit the audience’s emotions and beliefs. The rampant spread of fake news has the potential to sway public opinion, influence decision-making processes, and even disrupt social and political landscapes (Khattar et al. 2019). Consequently, the development of effective methods for fake news detection has become an urgent necessity.

Recent years have witnessed significant advancements in the field of fake news detection, driven by a combination of technological innovations and growing awareness about the potential consequences of misinformation (Zhou et al. 2020). This dynamic landscape poses both opportunities and challenges, as the creators of fake news constantly evolve their strategies to bypass detection mechanisms. From sophisticated AI-generated articles to meticulously doctored images and videos (Orhan 2023), the arsenal of fake news has expanded, demanding more sophisticated and adaptable detection techniques.

Several instances of fake news are depicted in Fig. 1. These fake news instances gained significant traction during the COVID-19 pandemic and the 2016 U.S. General Presidential Election (Kaliyar et al. 2021a).

This article delves into the latest trends and challenges surrounding fake news detection. It explores the evolving techniques employed by purveyors of misinformation and highlights the innovative strategies researchers and technologists have devised to counteract them. From natural language processing and machine learning algorithms to data mining and social network analysis, a multitude of approaches (Varghese et al. 2024) are being harnessed to differentiate between genuine news and deceptive content. However, amidst this progress, there remain formidable obstacles such as the lack of a universally agreed-upon

✉ Hemang Thakar
thakarhemang12@gmail.com

Brijesh Bhatt
brij.ce@ddu.ac.in

[1] Computer Science and Engineering, Charusat University, 139, CHARUSAT Campus, Highway, Off, Nadiad - Petlad Rd, Changa 388421, Gujarat, India

[2] Computer Engineering Department, Dharmsinh Desai University, College Rd, Chalali, Nadiad 387001, Gujarat, India

Published online: 30 August 2024

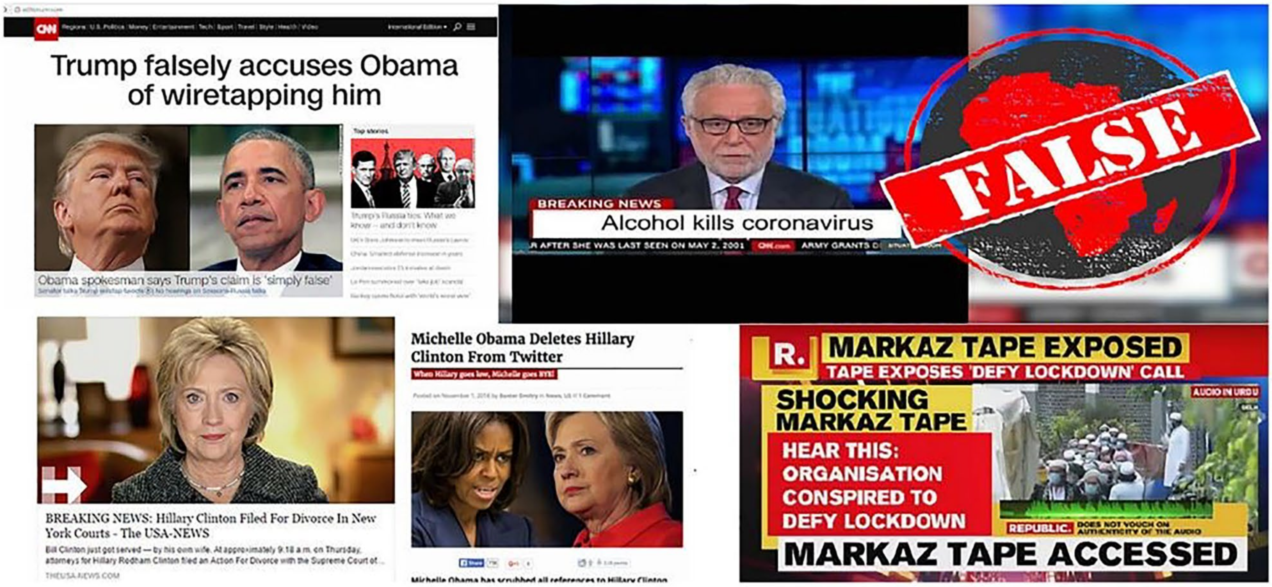
CNN
Trump falsely accuses Obama of wiretapping him
Obama spokesman says Trump's claim is 'simply false'
Top stories
Trump's Russia ties: What we know – and don't know
BREAKING NEWS
Alcohol kills coronavirus
R AFTER SHE WAS LAST SEEN ON MAY 2, 2001
ARMY GRANTS
FALSE
BREAKING NEWS: Hillary Clinton Filed For Divorce In New York Courts - The USA-NEWS
Bill Clinton just got served — by his own wife. At approximately 9:18 a.m. on Thursday, attorneys for Hillary Rodham Clinton filed an Action For Divorce with the Supreme Court of ...
THEUSA-NEWS.COM
Michelle Obama Deletes Hillary Clinton From Twitter
When Hillary goes low, Michelle goes BYE!
Michelle Obama has scrubbed all references to Hillary Clinton
R.
MARKAZ TAPE EXPOSED
TAPE EXPOSES 'DEFY LOCKDOWN' CALL
SHOCKING MARKAZ TAPE
HEAR THIS: ORGANISATION CONSPIRED TO DEFY LOCKDOWN
AUDIO IN URDU
REPUBLIC
DOES NOT VOUCH ON AUTHENTICITY OF THE AUDIO
MARKAZ TAPE ACCESSED

Fig. 1 Illustrations of misleading information circulated across social media. (Kaliyar et al. 2021a)

definition of fake news, the ethical implications of content moderation, and the balance between freedom of expression and the need to curb misinformation.

As we navigate this intricate landscape, it becomes crucial to understand the technical aspects of fake news detection and the broader societal and psychological factors that contribute to its dissemination and impact (Roy et al. 1811). By shedding light on the latest developments, this article aims to contribute to the ongoing discourse on combating fake news and promoting media literacy in an increasingly information-saturated world. As technology and deception continue intertwining, staying ahead in the battle against fake news requires (Bharadwaj and Shao 2019) vigilance, collaboration, and a multidimensional approach encompassing technology, psychology, and critical thinking.

Social media has transformed into a platform for sharing information, ideas, and feelings across the globe. However, this convenience has also facilitated the spread of misinformation, which can be disseminated quickly, cheaply, and maliciously. Spreading false information is often used to damage public information, organizations, and even countries, highlighting the importance of identifying misleading information (Oshikawa et al. 1811). Research is being done to create reliable and accurate algorithms that can automatically identify false information on social media to solve this issue. These automated applications are made using cutting-edge technology like data mining, machine learning, and natural language processing (NLP) (Shu et al. 2017).

1.1 Motivation and research objective

The field of fake news detection stands as a thriving research domain, drawing the keen interest of researchers worldwide. Significant scope for enhancement emerges within the realm of fake news detection, primarily due to the limited availability of context- specific news data for training purposes. The adoption of deep learning methodologies in fake news detection presents a distinctive advantage over conventional approaches, given their prowess in extracting advanced features from the data. These aforementioned challenges and opportunities serve as the driving force behind our endeavor to construct an efficient deep-learning model dedicated to the task of fake news detection.

1.2 Existing methodologies for the identification of fake news

Identifying fake news poses a significant challenge due to its deliberate intent to distort information. Preceding theories play a crucial role in directing investigations into counterfeit news detection, employing diverse classification models. Current insights into detecting fake news can be broadly grouped into two main categories: (i) Learning based on News Content, and (ii) Learning based on Social Context.

1.2.1 News content-based learning

News Content-based learning (Jain et al. 2022; Zhou et al. 2020; Dong et al. 2020; Sadeghi et al. 2022; Galli et al. 2022; Bra¸soveanu, A.M., Andonie, R. 2021; Verma et al. 2021; Shishah 2021) involves analyzing the textual and linguistic characteristics of news articles to distinguish between genuine information and fake news. This approach hinges on the understanding that deceptive content often exhibits linguistic anomalies, sensationalism, or lacks credible sources. By examining the structural attributes, writing style, and vocabulary usage within articles, machine learning algorithms can be trained to uncover patterns indicative of falsified information.

Through the utilization of Natural Language Processing (NLP) techniques, such as sentiment analysis, text summarization (Galli et al. 2022; Bra¸soveanu, A.M., Andonie, R. 2021; Reddy et al. 2020; Rani et al. 2022; Palani et al. 2022; Rai et al. 2022; Shan et al. 2021; Kaliyar et al. 2021b; Jarrahi and Safari 2023) and language models, this approach aims to identify inconsistencies, exaggerated claims, and linguistic markers commonly associated with misinformation. For instance, excessive use of emotional language, hyperbolic statements, or the absence of verifiable sources can raise red flags about the authenticity of the content.

Furthermore, this method is fortified by the accumulation of labeled datasets (Zhou et al. 2020; Palani et al. 2022; Rai et al. 2022; Jarrahi and Safari 2023) containing both genuine and fake news articles. Machine learning algorithms can then be trained on these datasets, allowing them to learn the nuanced distinctions between reliable and deceptive content. The process involves feature extraction, wherein relevant linguistic attributes are quantified, and classifiers are employed to differentiate between the two categories.

While News Content-based learning (Kaliyar et al. 2021a, 2021b; Sadeghi et al. 2022; Verma et al. 2021; Souza et al. 2022; Galende and Hern´andez-Pen˜aloza, G., Uribe, S., Garc´ıa, 2022; Nassif et al. 2022; Mughaid and Al-Zu'bi, S., Al Arjan, A., Al-Amrat, R., Alajmi, R., Zitar, R.A., Abualigah, L. 2022; Mohapatra et al. 2022) holds promise in detecting fake news, it is not without limitations. The constantly evolving nature of deceptive strategies demands ongoing updates to the algorithms. Additionally, this approach might struggle with subtle instances of misinformation that do not overtly deviate in language use or style. Balancing the need for accurate detection with potential false positives remains a challenge, as certain linguistic features might be shared between genuine news and well-crafted fake stories.

In essence, News Content-based learning (Li et al. 2021, 2020; Ying et al. 2021; Ma et al. 2015; Wang et al. 2021; Liu and Wu 2020) forms a pivotal part of the arsenal against fake news, leveraging linguistic and textual cues to unravel the threads of deception woven within the fabric of information. Its integration with other approaches, such as Social Context-based learning, holds the potential to enhance the accuracy and robustness of fake news detection systems.

1.2.2 Social context-based learning

Social Context-based learning (Li et al. 2021; Ying et al. 2021; Ma et al. 2015; Wang et al. 2021) involves analyzing the social interactions and dynamics surrounding news articles to assess their credibility and authenticity. This approach recognizes that the dissemination and reception of news are deeply intertwined with the social ecosystem in which they exist. By examining factors such as user engagement, sharing patterns, and the credibility of sources, this method aims to uncover signals that can help distinguish between genuine news and fake information. One of the key components of Social Context-based learning (Kaliyar et al. 2021a, 2021b; Sadeghi et al. 2022; Verma et al. 2021; Souza et al. 2022; Galende and Hern´andez-Pen˜aloza, G., Uribe, S., Garc´ıa, 2022; Nassif et al. 2022; Mughaid and Al-Zu'bi, S., Al Arjan, A., Al-Amrat, R., Alajmi, R., Zitar, R.A., Abualigah, L. 2022; Mohapatra et al. 2022) is the analysis of the propagation patterns of news articles across social media platforms. The rapid sharing of fake news often leads to its viral spread, driven by emotional responses and confirmation bias. By tracking the velocity and volume of shares, likes, comments, and retweets, algorithms can identify articles that are gaining traction unusually quickly or within specific echo chambers.

Furthermore, the credibility of the sources sharing the news plays a critical role. Social Context-based learning (Dong et al. 2020; Ying et al. 2021) involves assessing the authority and authenticity of the accounts sharing the information. Accounts with a history of sharing trustworthy content and a diverse range of sources are more likely to share accurate news. Conversely, accounts that predominantly share sensational or misleading information might raise suspicions.

Contextual analysis also contributes to this approach. Understanding the broader context in which a news article is shared, including the events and conversations surrounding it, can provide insights into its accuracy (Devlin et al. 1810; Reis et al. 2019; P´erez-Rosas, V., Kleinberg, B., Lefevre, A., Mihalcea, R. 1708). Additionally, detecting inconsistencies between a news article and verifiable facts can help identify potential misinformation.

Social Context-based learning is enhanced through the utilization of network analysis, sentiment analysis, and machine learning methodologies (Liu and Wu 2020; Long et al. 2017; Ozbay and Alatas 2021). By modeling the complex relationships between users, content, and interactions, algorithms can learn to differentiate between genuine news and fake news based on social dynamics. However, challenges exist in this approach as well (Liu and Wu 2020; Li et al. 2020). Misinformation campaigns can manipulate social dynamics, employing tactics to artificially inflate engagement metrics. Moreover, relying solely on social context might not catch sophisticated fake news stories that avoid triggering suspicious patterns.

In conclusion, Social Context-based learning complements (Kaliyar et al. 2021a, 2021b; Nassif et al. 2022; Mughaid and Al-Zu'bi, S., Al Arjan, A., Al-Amrat, R., Alajmi, R., Zitar, R.A., Abualigah, L. 2022; Mohapatra et al. 2022) other fake news detection strategies by tapping into the intricate web of social interactions and human behaviors. Its ability to uncover anomalies in sharing patterns and evaluate the credibility of sources offers a valuable perspective in the ongoing battle against the spread of fake news. When integrated with News Content-based learning and other approaches, it contributes to a more comprehensive and effective fake news detection framework.

1.2.3 Hybrid models

Hybrid models combine both content-based and context-based approaches to leverage the strengths of each methodology (Comito et al. 2023). These models can provide a more comprehensive analysis by considering both the textual content and the contextual information.

2 Recent Advancements

1. **Content and social context fusion** presented a hybrid model that fuses content-based features with social context features (Orhan 2023). This model uses a multimodal neural network that simultaneously processes textual content using BERT and social context using GNNs (Galende and Hern´andez-Pen˜aloza, G., Uribe, S., Garc´ıa, 2022). The fusion layer combines these features to improve detection accuracy, particularly in cases where either content or context alone is insufficient.
2. **Multi-view learning** proposed a multi-view learning framework that incorporates multiple perspectives, including content (Galli et al. 2022), user behavior, and propagation patterns. By using attention mechanisms to weigh the importance of each view dynamically, the model can adapt to different types of fake news scenarios, enhancing its robustness.

These models represent cutting-edge techniques in fake news detection, combining advancements in NLP and network analysis to address the complex challenge of identifying fake news on social media platforms (Khalil et al. 2024). Integrating recent research findings into your paper will provide a comprehensive overview (TS, S.M., Sreeja, P. 2024) of the current state-of-the-art in this field.

2.1 Characteristics of fake news detection

Fake news detection is a critical area of research, focusing on identifying false or misleading information disseminated through various media channels, especially social media (Rastogi and Bansal 2023). This task involves several distinct characteristics that researchers aim to address. First, fake news often exhibits sensationalist language and exaggerated claims intended to elicit strong emotional reactions from readers, making it important for detection systems to analyze linguistic features. Second, the context in which the news appears is crucial; understanding the source, author, and the spread pattern on social networks helps in evaluating the credibility of the information. Third, fake news frequently leverages multimedia elements like images and videos, requiring advanced detection systems to integrate textual analysis with image and video verification (Athira et al. 2023). Fourth, the temporal aspect is significant, as fake news can spread rapidly, necessitating real-time or near-real-time detection capabilities. Additionally, adversarial techniques (Akdag and Cicekli 2024) are used to bypass detection mechanisms, highlighting the need for robust, adaptive models that can counteract these efforts. Hybrid models, which combine content-based and context-based features, are emerging as effective solutions, offering improved accuracy and resilience against sophisticated fake news tactics (Ozbay and Alatas 2021; Shu et al. 2019). Ultimately, the goal is to develop comprehensive systems that can accurately identify fake news, mitigate its spread, and enhance public trust in information.

2.2 Supervised fake news detection

Supervised fake news detection involves training models using labeled datasets where each news item is pre-annotated as either fake or real. This method relies heavily on the availability of large, accurately labeled datasets, which serve as ground truth for the learning algorithms. Supervised models, such as Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and Long Short-Term Memory networks (LSTM), leverage these dataset (Athira et al. 2023) to learn the distinguishing features of fake news, including linguistic patterns, semantic content, and contextual cues. The effectiveness of supervised learning (Kumar and Taylor 2024) is often high when the training data is comprehensive and representative of the variations in news content. However, obtaining such high-quality labeled data is challenging, time-consuming, and expensive, which limits the scalability of supervised approaches.

2.3 Weakly supervised fake news detection

On the other hand, weakly supervised fake news detection aims to alleviate the dependency on extensive labeled datasets by utilizing partially labeled or noisy data. This approach (Rastogi and Bansal 2023) leverages various strategies, such as semi-supervised learning, where a small amount of labeled data is used in conjunction with a larger pool of unlabeled data, and transfer learning, where models pretrained on related tasks are fine-tuned on the target task. Weakly supervised methods also include distant supervision, where external knowledge sources like fact-checking websites provide weak labels. These approaches enable models to learn effectively from less-than-perfect data, making them more adaptable and scalable. Weakly supervised models (Akdag and Cicekli 2024) are particularly useful in dynamic environments like social media, where new and diverse fake news content emerges rapidly. Despite their flexibility, these models (Zhao et al. 2306) may struggle with accuracy and reliability compared to fully supervised models, necessitating ongoing refinement and validation.

3 Related work

Our study sought to bridge the knowledge gap in the area by offering a comprehensive review of the current methods for detecting fake news and promoting multidisciplinary research collaboration. The primary goal of this paper is to provide an overview of the current state of research on the topic.

To achieve this goal, we conducted a thorough review of various solutions that are currently being used to detect fake news. We analyzed the use of machine learning models, network propagation models, and fact-checking methodologies for detecting fake news. In particular, our study focused on how researchers develop and use machine learning models to identify and classify fake news, as well as the tools they employ for this purpose. Furthermore, we also discussed the research challenges that are still open in this field.

The paper (Ahmed et al. 2018) presents a novel n-gram model that automatically detects false information, with a particular focus on fake news and misleading judgments. The study employs two distinct attribute abstraction methods and six different machine-learning classification algorithms. Prepossessing involves removing stop words and stemming keywords to identify misleading information effectively. The classifier is trained using two feature extraction techniques, TF and TF-IDF, in the final classification stage. The research evaluates six machine learning algorithms: SGD, SVM (Sadeghi et al. 2022; Verma et al. 2021; Reddy et al. 2020; Palani et al. 2022; Shan et al. 2021), LSVM (Kaliyar et al. 2021a; Shishah 2021; Rai et al. 2022), KNN (Sadeghi et al. 2022; Verma et al. 2021), LR(Sadeghi et al. 2022; Rani et al. 2022), and DT(Kaliyar et al. 2021a; Verma et al. 2021).

The paper examines (Hirlekar and Kumar 2020) the current state of research on fake news and proposes theoretical and practical approaches to categorize and intervene in this problem. Text mining is one such machine learning approach used to detect fake news, hoaxes, and misinformation. The study proposes a complex classification scheme, including neural networks that use traditional classification procedures. The report recommends identifying fake news using essential text qualities that can be produced independently of platform and language (Faustini and Covoes 2020).

The study compares five datasets, which include articles and posts from social media in three distinct categories of languages, to standards and finds favorable outcomes. The study also examines how training factors affect other common natural language processing algorithms like Word2Vec and bag-of-words.

The study proposes a hybrid attention LSTM model and uses the Wang (Wang 1705) LIAR dataset (Jain et al. 2022; Dong et al. 2020; Sadeghi et al. 2022; Galli et al. 2022) from PolitiFact, which has subject, text, and speaker profiles for 12,836 news items from 3,341 speakers. The results show that the model outperforms recent reference dataset-based models by 14.5%. This demonstrates the importance of the speaker's profile in determining news trustworthiness (Galli et al. 2022).

The paper begins with an examination of the need for automatic fake news detection, comparing and discussing various techniques' findings on the most critical new standard datasets. The research focuses on the LIAR (Jain et al. 2022; Dong et al. 2020; Sadeghi et al. 2022; Galli et al. 2022; Bra¸soveanu, A.M., Andonie, R. 2021; Comito et al. 2023), FEVER, and FAKENEWSNET (Sadeghi et al. 2022; Verma et al. 2021; Souza et al. 2022; Galende and Hern´andez-Pen˜aloza, G., Uribe, S., Garc´ıa, 2022; Nassif et al. 2022; Mughaid and Al-Zu'bi, S., Al Arjan, A., Al-Amrat, R., Alajmi, R., Zitar, R.A., Abualigah, L. 2022) datasets, with the LIAR dataset showing excellent accuracy with LSTM and attention LSTM-based models. The FEVER dataset also uses an attention-based LSTM-based model (Kaliyar et al. 2021a; Shishah 2021; Rai et al. 2022) to achieve outstanding accuracy, and the GCN-based model using the FAKENEWSNET dataset achieves complete accuracy (Oshikawa et al. 1811).

Research on how to identify fake news has been ongoing, and numerous algorithms have been created to do so. To detect fake news, researchers have used a variety of models, including convolutional neural networks, long-short-term memory networks, and bidirectional LSTM (Comito et al.

2023; Li et al. 2022). They obtained word vector representations using glove, an unsupervised machine learning approach, and then modeled a deep neural network using CNN and Max-pooling. After that, the gradient disappeared and issues with long-term dependency were removed using Bi-LSTM (Sadeghi et al. 2022; Rani et al. 2022; Mughaid and Al-Zu'bi, S., Al Arjan, A., Al-Amrat, R., Alajmi, R., Zitar, R.A., Abualigah, L. 2022). The Attention Mechanism, which has been effective in various tasks such as machine translation and picture captioning, was used, and a dropout layer was used in the last phase to prevent overfitting. The researchers achieved a 71.2% accuracy rate for the approved testing dataset (Rani et al. 2022).

In another study, the authors categorized the critical factors for each job to investigate the use of multiple supervised learning classifiers, such as KNN(Sadeghi et al. 2022; Verma et al. 2021; Reddy et al. 2020), NB(Sadeghi et al. 2022; Verma et al. 2021; Reddy et al. 2020; Rani et al. 2022; Palani et al. 2022), RF, SVM(Sadeghi et al. 2022; Verma et al. 2021; Reddy et al. 2020; Palani et al. 2022; Shan et al. 2021), and XGB, and the accuracy and F1 score obtained by each classifier. RF and XGB outperformed other classifiers, and it was found that distinguishing fake from genuine articles on a significant, recently accessible, and wholly labeled dataset was tough. They also discussed how supervised learning models could help fact-checkers analyze digital data and come up with solid conclusions (Verma et al. 2021).

To detect fake news, another research employed semantic characteristics and text mining and compared RNN to a naive Bayes classifier and random forest classifier using different groups of linguistic features. Random forest outperformed Naive Bayes in trials when utilizing different features, with a result of 95.66%. The researchers used a Kaggle real-or-fake news dataset for their experiments (Bharadwaj and Shao 2019).

In yet another study, the authors proposed a Multi-source, Multi-class Fake News Detection system that combines convolutional neural networks to analyze the local structure of every word in a statement, LSTM (Kaliyar et al. 2021a; Shishah 2021; Rai et al. 2022) to analyze temporal relationships across the text, and an integrated network to concatenate the last hidden outputs. This technique combines the best characteristics of both systems, as LSTM performs better with lengthier jail sentences (Karimi et al. 2018).

The authors of another study proposed a novel method for detecting fake news at the KE level, which entails representing the claims in the news item as a multimedia knowledge graph and recognizing the misleading aspects in the kind of KEs for a high degree of explainability. They developed a logically structured approach called InfoSurgeon for detecting disinformation in news articles that involve source context, semantic representation, multimedia information components, and previous knowledge. They also proposed a new benchmark for identifying fake news at the KE level via a silver standard annotation dataset (15,000 multimedia article pairs) generated automatically using KG-influenced natural language generation (Fung et al. 2021).

The authors present tools and models that aim to address the challenge of detecting fake news and supporting their study. They created two new datasets, spanning seven different fields, using a combination of human and crowdsourced annotations and data directly obtained from the internet (Monti et al. 1902; Hu et al. 2022). Exploratory tests were conducted using these datasets to identify linguistic features that could potentially be indicative of fraudulent content. Based on these features, the authors developed fake news detectors that achieved up to 78% precision. To provide context for their findings, they compared the efficacy of their detection systems to an objective human baseline (P´erez-Rosas, V., Kleinberg, B., Lefevre, A., Mihalcea, R. 1708).

Table 1 Model references

References	Model
Sadeghi et al. (2022), Bra¸soveanu and Andonie (2021), Rani et al. (2022)	LR
Sadeghi et al. (2022), Bra¸soveanu and Andonie (2021), Verma et al. (2021), Reddy et al. (2020), Palani et al. (2022), Shan et al. (2021), Ma et al. (2015)	SVM
Kaliyar et al. (2021a), Sadeghi et al. (2022), Verma et al. (2021), Reddy et al. (2020), Rani et al. (2022)	KNN
Kaliyar et al. (2021a), Bra¸soveanu and Andonie (2021), Verma et al. (2021), Ma et al. (2015)	DT
Kaliyar et al. (2021a), Bra¸soveanu and Andonie (2021), Verma et al. (2021), Reddy et al. (2020), Rani et al. (2022), Palani et al. (2022), Ma et al. (2015)	RF
Sadeghi et al. (2022), Verma et al. (2021), Reddy et al. (2020), Rani et al. (2022), Palani et al. (2022)	NB
Kaliyar et al. (2021a), Shishah (2021), Rai et al. (2022)	LSTM
Sadeghi et al. (2022), Galli et al. (2022), Bra¸soveanu and Andonie (2021), Rani et al. (2022), Mughaid et al. (2022), Mohapatra et al. (2022)	BiLSTM
Kaliyar et al. (2021a), Dong et al. (2020), Galli et al. (2022), Bra¸soveanu and Andonie (2021), Verma et al. (2021)	CNN
Jain et al. (2022), Kaliyar et al. (2021a), Sadeghi et al. (2022); Galli et al. (2022); Verma et al. (2021), Shishah (2021), Palani et al. (2022), Rai et al. (2022), Shan et al. (2021), Nassif et al. (2022)	BERT

Table 1 is an indispensable resource for researchers and practitioners in the field of data analysis and research. This table meticulously catalogs a wide range of datasets, serving as a comprehensive reference guide. Whether one is delving into machine learning, statistical analysis, or any data-driven investigation, Table 1 simplifies the process of dataset selection and comparison. It embodies the foundational principle that robust research relies on the quality and relevance of data, making it an invaluable asset in the pursuit of knowledge and innovation across various domains.

Table 2 serves as a valuable resource for gaining insights into the extensive body of research dedicated to understanding and addressing issues related to bots, clickbaits, rumors, and the analysis of content and context in digital communication. This table presents a consolidated view of the diverse studies, methodologies, and findings within this domain, offering a comprehensive snapshot of the collective efforts made by researchers and scholars. It not only highlights the breadth and depth of research but also provides a convenient reference point for those seeking to explore specific topics within this multifaceted field. In an era marked by the rapid dissemination of information through digital channels, Table 3 plays a pivotal role in promoting informed decision-making and fostering a deeper understanding of the complexities surrounding online content and its impact on society (Fig. 2).

4 Different types of fake news detection Model

4.1 Machine learning technique

Initially, machine learning algorithms were developed to detect fake news due to the belief that it is created for financial and political gain (Faustini and Covoes 2020). Since fake news often includes persuasive and argumentative language, the retrieval of written text and linguistic elements is required for machine learning. The author utilized the Naive Bayes classifier to recognize linguistic features such

Table 2 Dataset references

References	Dataset
Jain et al. (2022), Roy et al. (2018), Dong et al. (2020), Sadeghi et al. (2022), Galli et al. (2022), Bra¸soveanu and Andonie (2021), Jarrahi and Safari (2023), Galende et al. (2022), Long et al. (2017), Ozbay and Alatas (2021), Wang (1705), Rajalaxmi et al. (2022), Truicˇa and Apostol (2023)	LIAR
Zhou et al. (2020), Galli et al. (2022), Bra¸soveanu and Andonie (2021), Verma et al. (2021), Shishah (2021), Reddy et al. (2020), Rani et al. (2022), Palani et al. (2022); Rai et al. (2022), Shan et al. (2021), Kaliyar et al. (2021b), Jarrahi and Safari (2023)	PolitiFact
Kaliyar et al. (2021a), Sadeghi et al. (2022), Verma et al. (2021), Kaliyar et al. (2021b), Souza et al. (2022), Galende et al. (2022), Nassif et al. (2022), Mughaid et al. (2022) and Mohapatra et al. (2022)	FakeNewsNet
Verma et al. (2021), Kaliyar et al. (2021b), Ozbay and Alatas (2021), Steni Mol and Sreeja (2024), Truicˇa and Apostol (2023), Malhotra and Malik (2024), Katarya et al. (2022)	BUZZFEED
Zhou et al. (2020), Galli et al. (2022), Palani et al. (2022), Rai et al. (2022), Jarrahi and Safari (2023), Steni Mol and Sreeja (2024), Katarya et al. (2022)	GossipCop
Khattar et al. (2019), Li et al. (2021), Ying et al. (2021), Ma et al. (2015), Wang et al. (2021), Liu and Wu (2020), Li et al. (2020), Li et al. (2022), Katarya et al. (2022), Wang et al. (2018)	Weibo
Dong et al. (2020), Galli et al. (2022), Ying et al. (2021), Li et al. (2022) and Katarya et al. (2022)	PHEME

Table 3 Comprehensive overview of research utilized in bots, clickbaits, rumors, content, and context analysis

Reference	Approach	Size of dataset	Elevation Metrics	Platform
Verma et al. (2021)	NB	501 fraudulent accounts and 499 legitimate accounts	ROC curve, F1 score and confusion matrix	Twitter
Reddy et al. (2020)	SVM	54 million users, 1.9 billion links, 1.8 billion tweets	Precision, recall, Micro F1 and Macro F-1	Twitter
Bra¸soveanu and Andonie (2021)	Bayesian classifier	25,000 users, 500,000 tweets, and 49 million followers	Precision	Twitter
Sadeghi et al. (2022)	LR	300 rumors, 2595 news articles, headlines	Accuracy, precision and recall	Twitter
Shishah (2021)	LSTM-RNN	3,600 fake news articles and 68,892 real news articles	Micro-F1 and Macro-F1	Twitter
Dong et al. (2020)	Hybrid CNN	12,836 concise statements	Accuracy	Politifact
Ma et al. (2015)	RF	1,627 articles	Accuracy, precision, recall and F1	Facebook

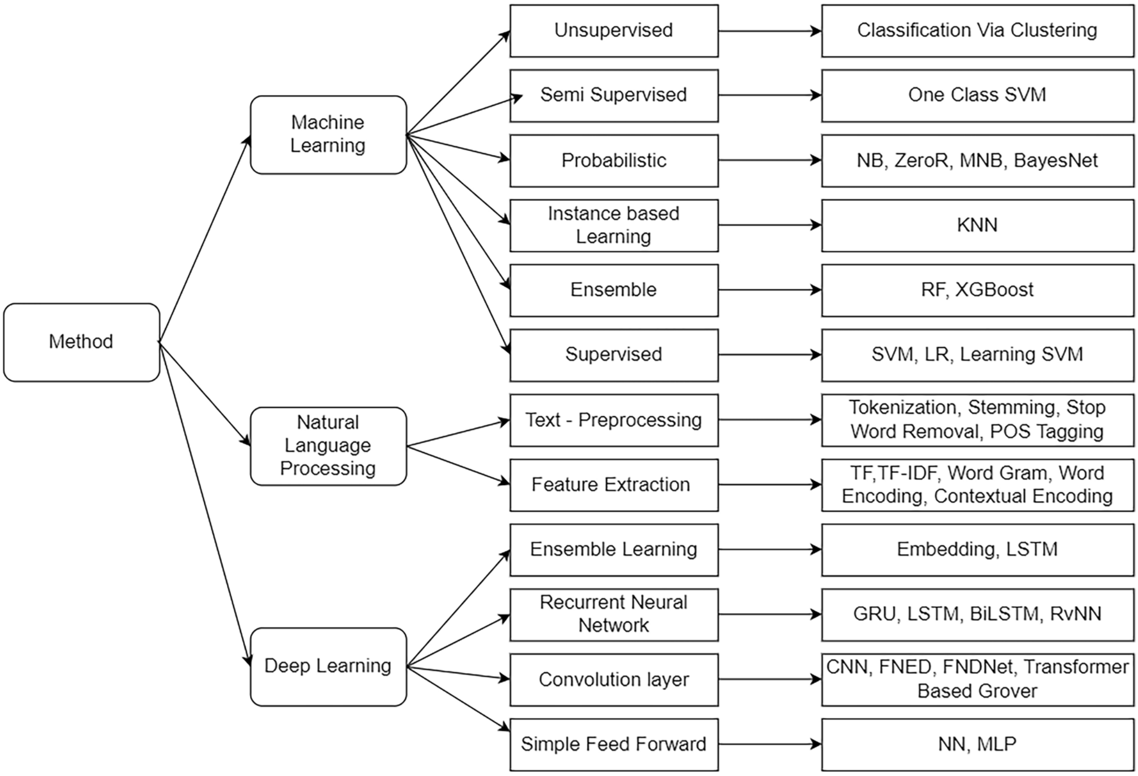

Fig. 2 Techniques Utilized for Detecting False Information

as vocabulary, word count, length, and grammatical style, including text summarization and characterization (Oliveira et al. 2020). However, some fake news categories, such as clickbait articles, have high click-through rates due to their alluring nature, which cannot be detected by this technology.

The authors suggest a machine learning model as a solution, employing gradient- boosted decision trees to identify fake news effectively, resulting in high classification accuracy. They have pinpointed the significance of a casual tone in the creation of clickbait articles. (Elhadad et al. 2020).

Additionally, the author has devised a machine learning model capable of discerning and forecasting whether an article qualifies as clickbait, leveraging features such as URL, content, and title. They used Yahoo aggregate data to construct a training set of 1349 clickbait URLs and a testing set of 2724 non-clickbait URLs. The author categorized eight types of clickbait, such as exaggeration, tease, confusion, provocation, formatting, bait-and-switch, graphic, and wrong, to identify spam mail and websites (Faustini and Covoes 2020; Silva et al. 2020) (Fig. 3).

4.2 Natural language processing technique

Natural language processing (NLP) has become a valuable tool in detecting fraud through a variety of techniques, including grammatical and syntactic analysis, correlation, clustering, and boolean text classification, which identifies news as true or false. When detection is challenging, a third category may be introduced to differentiate between temporary actual and temporary fake situations. Utilizing the Text Segmentation method along with the Natural Language Toolkit, the Sentiment Score is subsequently calculated by examining carefully chosen and structured text for indications of fraudulent activity. In NLP, features like text quality and context are essential for accurate detection (Liu and Wu 2020). Stanford parser, a language, and syntactic analyzer claims to produce reliable results. Reality-proof (Silva et al. 2020) studies have shown that NLP is more effective than social authentication. The main aim is to identify syntactic and verbal cues that reveal linguistic disparities between individuals who tell lies and those who tell the truth.

4.3 Deep learning technique

Detecting different types of fake news in the context of deep learning involves the application of various machine-learning techniques to combat the proliferation of misinformation in today's digital age. The multifaceted nature of fake news necessitates a diverse range of approaches. Text-based fake news detection leverages recurrent neural networks (RNNs) and convolutional neural networks (CNNs) to scrutinize linguistic patterns, sentiment, and contextual cues in textual

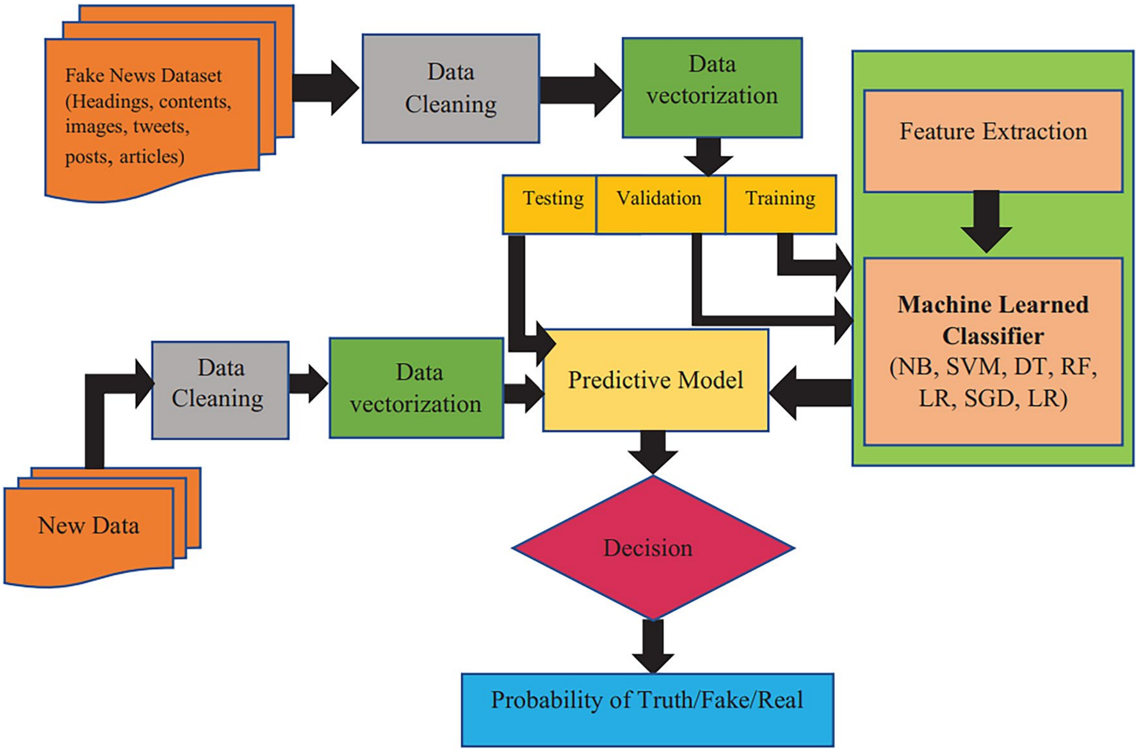
Fake News Dataset (Headings, contents, images, tweets, posts, articles)
Data Cleaning
Data vectorization
Testing
Validation
Training
Feature Extraction
Machine Learned Classifier (NB, SVM, DT, RF, LR, SGD, LR)
New Data
Data Cleaning
Data vectorization
Predictive Model
Decision
Probability of Truth/Fake/Real

Fig. 3 Machine learning architecture of fake information detection. (Meel and Vishwakarma 2020)

content. These models analyze the language used in news articles, social media posts, and other textual sources to identify deceptive narratives. Image-based fake news detection, on the other hand, harnesses deep convolutional neural networks (CNNs) to scrutinize visual elements within images. These models can uncover alterations, forgeries, or inconsistencies in visual content, a vital component of debunking photo-manipulated stories. Audio-based fake news detection delves into audio files using models like CNNs (Liu and Wu 2020) and RNNs to identify voice impersonations, audio tampering, or anomalies in spoken content. Combining text, images, and audio, multi-modal fake news detection employs advanced architectures such as Transformer-based models and multi-modal neural networks (MNNs) (Karimi et al. 2018) to fuse and analyze data from various media sources. These models provide a more comprehensive evaluation of information credibility by considering the collective impact of multiple modalities. Overall, the use of deep learning in these various modes empowers researchers and developers to stay ahead in the battle against fake news(Bharadwaj and Shao 2019), offering a multi-pronged approach to safeguard the accuracy and integrity of information in an increasingly digital and interconnected world (Fig. 4).

Tables 4 and 5 offer a comprehensive overview of ensemble-based machine/deep learning methods that have achieved evaluation metric scores surpassing 90%. The highlighted

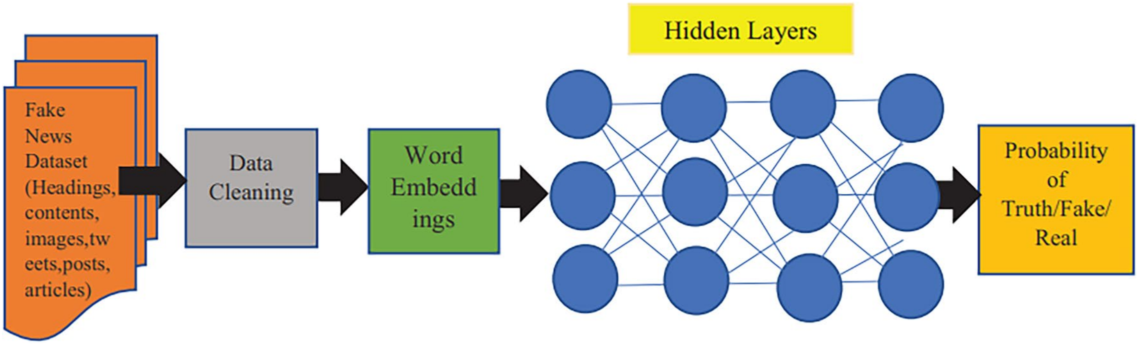
Hidden Layers
Fake News Dataset (Headings, contents, images,tweets,posts, articles)
Data Cleaning
Word Embeddings
Probability of Truth/Fake/ Real

Fig. 4 Deep learning architecture of fake information detection. (Meel and Vishwakarma 2020)

Table 4 Metrics assessing performance for the most effective approach in the previously cited research using ML approaches

Reference	ML Approaches		
	Accuracy	Precision	F1 score
	Above 90% and Model	Above 90% and Model	Above 90% and Model
Faustini and Covoes (2020)	Yes and SVM	No	Yes and SVM
Silva et al. (2020)	Yes and SVM,and LR,RF	Yes and SVM, and LR,RF	Yes and SVM,and LR,RF
Oliveira et al. (2020)	No	Yes and LSA	No
Elhadad et al. (2020)	Yes and DT,and NN, LR	Yes and DT,and NN,LR	Yes and DT,and NN, LR
Kausar et al. (2020)	Yes and CNN	Yes and CNN	Yes and CNN

Table 5 Metrics assessing performance for the most effective approach in the previously cited research using DL approaches

Reference	DL Approaches		
	Accuracy	Precision	F1 score
	Above 90% and Model	Above 90% and Model	Above 90% and Model
Wang et al. (2021)	Yes and CNN	Yes and CNN	Yes and CNN
Umer et al. (2020)	Yes and CNN, LSTM	Yes and CNN, LSTM	Yes and CNN, LSTM
Huang and Chen (2020)	Yes and LSTM	Yes and LSTM	Yes and LSTM
Kaliyar et al. (2020)	Yes and FNDNet	Yes and FNDNet	Yes and FNDNet
Li et al. (2020)	Yes- and MCNN	No	Yes and MCNN

cells spotlight the researchers' top accomplishments in these tables.

5 Dataset

BuzzFeed (Santia and Williams, J.: Buzzface 2018) Comprising a full news sample published on Facebook, this dataset encompasses content from 9 news agencies spanning September 19 to 23, as well as September 26 and 27, a week preceding the 2016 U.S (Hu et al. 2022). Election. Each post, alongside its associated articles, underwent validation by five BuzzFeed journalists individually. The dataset encompasses a total of 1627 articles, with a breakdown of 826 mainstream articles, 356 left-wing articles, and 545 right-wing articles.

LIAR (Wang 1705) encompasses a collection of 12,836 real-world news articles curated from PolitiFact, each piece of news in this dataset is categorized based on a six-grade truthfulness (Wang 1705; Hu et al. 2022) scale: true, false, half-true, part-true, barely-true, and mostly- true. Additionally, the dataset provides supplementary details about the subjects covered, political affiliations, contextual information, and speakers mentioned in the news articles.

Wibo (Li et al. 2021) is presenting a multi-domain fabricated news dataset in the Chinese language, each dataset entry is accompanied by an annotated domain label. The dataset encompasses fabricated as well as authentic news articles sourced from Sina Weibo, spanning the period from December 2014 to March 2021. Regarding fabricated content, the Weibo21 dataset comprises news articles officially identified as misinformation by the Weibo Community Management Center.

PolitiFact (Zhou et al. 2020) is a renowned nonprofit fact-checking website that operates within the United States and specializes in evaluating political statements and reports. The dataset from PolitiFact encompasses news articles published between May 2002 and July 2018. Verified by domain experts, the dataset includes definitive labels (false or true) assigned to the news content. The content in the PolitiFact dataset primarily consists of statements or news articles disseminated by political figures (such as Congress members, White House staff, and lobbyists) and political groups, all of which have undergone thorough fact-checking by PolitiFact.

GossipCop (Zhou et al. 2020) is operating as a fact-checking platform, the GossipCop dataset pertains to news articles released between July 2000 and December 2018. The dataset incorporates domain experts who meticulously assign definitive labels to news content, thereby upholding the accuracy and reliability of the news tags.

FakeNewsNet (Verma et al. 2021) is incorporating data sourced from the fact-checking platforms BuzzFeed and PolitiFact, this dataset encompasses news articles along with associated user details and retweet information. The dataset aggregates a combined total of 23,196 news articles and 69,733 retweets.

PHEME (Zubiaga et al. 2017) is a compilation consisting of tweets originating from the Twitter platform. Furthermore, the data was gathered from five distinct sources specializing in breaking news, with each source contributing a set of tweets. Each tweet within the dataset comprises both textual content and accompanying images.

We provide a summary of the publicly available datasets utilized, as depicted in Table 6. These datasets comprise data sourced from platforms such as Sina Weibo, Twitter, and various other social media platforms, along with information from fact- checking websites like BuzzFeedWeb, LIAR, and FakeNewsNet.

Table 7 presents a comprehensive overview of the results obtained from various associated models employed in the detection of fake news. This table offers valuable insights into the performance and effectiveness of different approaches and methodologies used to tackle the challenging task of identifying deceptive or misleading information in news sources. By summarizing the results from these associated models, Table 7 serves as a valuable reference for researchers, policymakers, and practitioners striving to enhance our understanding of fake news detection and develop more robust solutions to address this pressing issue in today's information landscape (Tables 8 and 9,10,11,12,13).

6 Performance measure formula

The analysis of all the gathered articles reveals that, in every case, one or more of the ten performance metrics, as depicted in Table 14, have been employed to assess the simulation outcomes. These metrics serve as indicators of a method's ability to detect. Additionally, Table 14 includes the formulas for each performance metric.

As indicated in the presented Table 14, a comprehensive evaluation of classifier performance is conducted across multiple dimensions, including Accuracy, Error Rate, Precision, Sensitivity, F1-Score, Specificity, Area Under the Curve, Geometric Mean, Miss Rate, False Discovery Rate, and Fall-Out Rate.

Within Table 14, the designations TP(True Positive) and TN(True Negative) refer to the count of accurately classified positive and negative instances respectively, while FP(False Positive) and FN(False Negative) refer to the count of positive and negatively labeled instances that were inaccurately classified.

The experimental outcomes stemming from the constructed models were rigorously assessed utilizing all the metrics enumerated in Table 14. The intention was to gauge the performance of distinct detection models from various vantage points rather than relying solely on a singular perspective.

7 Open issues and future research

Fake news has emerged as a significant challenge in the modern information landscape, fueled by the rapid dissemination of information through digital platforms and social media. While substantial progress has been made in developing fake news detection techniques, several open issues and

Table 6 Overview of fake news detection datasets

Reference	News context				Social context		
	Dataset	Text	Visual	User profile	Repost and response	Network	Lables
Jain et al. (2022; Dong et al. (2020; Sadeghi et al. (2022; Galli et al. (2022; Bra¸soveanu and Andonie (2021)	LIAR	Yes					6
Zhou et al. (2020; Galli et al. (2022; Bra¸soveanu and Andonie (2021; Verma et al. 2021), Shishah 2021; Reddy et al. 2020; Rani et al. 2022; Palani et al. 2022; Rai et al. 2022; Shan et al. 2021; Kaliyar et al. 2021b; Jarrahi and Safari 2023)	PolitiFact	Yes			Yes		2
Zhou et al. 2020; Palani et al. 2022; Rai et al. 2022; Jarrahi and Safari 2023)	GossipCop	Yes					2
Kaliyar et al. 2021a; Sadeghi et al. 2022; Verma et al. 2021; Kaliyar et al. 2021b; Souza et al. 2022; Galende and Hern´andez-Pen˜aloza, G., Uribe, S., Garc´ıa, 2022; Nassif et al. 2022; Mughaid and Al-Zu'bi, S., Al Arjan, A., Al-Amrat, R., Alajmi, R., Zitar, R.A., Abualigah, L. 2022; Mohapatra et al. 2022)	FakeNewsNet	Yes	Yes	Yes	Yes	Yes	2
Verma et al. 2021; Kaliyar et al. 2021b)	BUZZFEED	Yes					4
Li et al. 2021; Ying et al. 2021; Ma et al. 2015; Wang et al. 2021; Liu and Wu 2020; Li et al. 2020)	Weibo	Yes	Yes	Yes	Yes	Yes	2
Dong et al. 2020; Ying et al. 2021)	PHEME	Yes	Yes	Yes	Yes		2

Table 7 Overview of the results from associated models for detecting fake news

Reference	Model	Dataset	Accuracy (%)	Limitation	Description
Jain et al. 2022)	BERT	LIAR	46.36	A limitation of this paper is the focus on text information and the absence of a real-time detection system for analyzing visual content, which may not address the full spectrum of fake news dissemination	The authors utilize BERT to boost fake news detection through its contextual understanding and attention mechanisms, with a focus on accuracy improvement and future exploration of multi-modal and hybrid approaches
Zhou et al. 2020)	Bi-LSTM	Politifact	99.63	While the paper acknowledges its limitations, it falls short in providing explicit details about the present work's draw- backs and a clear road-map for addressing them in future research	Authors used Bi-LSTM to monitor incoming messages, detecting rumors in an integrated approach
Mohapatra et al. 2022)	Hybrid BiLSTM and self-attention model	FNN	98.65	A major limitation of the paper is its reliance on an imbalanced dataset, which may affect its applicability for multi-class classification, potentially leading to biased results	To effectively classify genuine and fake news articles by leveraging advanced pre-processing, word embedding, and layer combinations, achieving superior accuracy and the potential to enhance user reading experiences by reducing the presence of fake news
Palani et al. 2022)	CB-Fake	Gossipcop	91	A limitation of the paper is its focus on English fake news datasets, which may not fully generalize to other languages and domains, limiting its applicability	To address the limitations of existing fake news detection models, particularly in extracting informative features from both the textual and visual content of news articles, achieving superior performance compared to current state-of-the-art methods
Dong et al. 2020)	CNN-based SSL	LIAR and PHEME	83.36 and 60.64	A limitation of this paper is that the proposed method relies on a substantial amount of unlabeled data, which may not be readily available in practical situations, potentially limiting its immediate applicability	To the challenge of timely detecting fast-propagating fake news with limited labeled data, the model was able to effectively detect fake news using both labeled and unlabeled data from PHEME and LIAR datasets

Table 8 Performance of various models on the LIAR dataset

Reference	Dataset	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Sadeghi et al. 2022)		BiLSTM	41.31	–	–	–
Sadeghi et al. 2022)		BiGRU	40.13	–	–	–
Ozbay and Alatas 2021)		ASSO-OSIW	41	71.3	61.2	75.9
Jarrahi and Safari 2023)		SLCNN	33	–	–	–
Jain et al. 2022)		ELMo-enabled Attention-based Model	46.36	–	–	–
–	LIAR					
Wang 1705)		Hybrid CNN	27.4	–	–	–
Long et al. 2017)		Hybrid LSTM	41.5	–	–	–
Roy et al. 1811)		Bi–LSTM	42.65	–	–	–
Roy et al. 1811)		A Deep Ensemble Model (RNN-CNN)	44.87	–	–	–
Galli et al. 2022)		B.E.R.T	63	59.5	60.6	62.8
Bra¸soveanu, A.M., Andonie, R. 2021)		CapsNet	64.9	–	–	–
Rajalaxmi et al. 2022)		Optimized LSTM	45.23	–	–	–
Truic˘a, C.-O., Apostol, E.-S. 2023)		MisRoBÆRTa	24.62	–	–	–

Table 9 Performance of various models on the BuzzFeed dataset

Reference	Dataset	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Ozbay and Alatas 2021)		ASSO-OSIW	99.1	98.2	1.00	96.4
Truic˘a, C.-O., Apostol, E.-S. 2023)		SVM	78	–	–	–
TS, S.M., Sreeja, P. 2024)		Deep Bi-LSTM	92.8	–	–	–
–	BuzzFeed					
[Inline Image Removed]Malhotra, P., Malik, S.K. 2024)		CSSLnO-based DQNN	90.02	91.9	89.04	91.46
Katarya et al. 2022)		MSVM	94	–	–	–

Table 10 Performance of various models on the Politifact dataset

Reference	Dataset	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Jarrahi and Safari 2023)		SLCNN	99	97.9	1.00	98.8
Zhou et al. 2020)		att–RNN	76.9	73.5	94.2	82.6
Galli et al. 2022)		BERT	62.7	61.3	50.6	62.8
Bra¸soveanu, A.M., Andonie, R. 2021)		CapsNet	52.4	–	–	–
–	Politifact					
Varghese et al. 2024)		Distilled BERT	55	–	–	–
Kiruthika and Rajagopalan 2022)		hybrid LSTM– SVM	90	98	–	90
Palani et al. 2022)		CB–Fake	93	92	91	92
Katarya et al. 2022)		MSVM	98.7	–	–	–

avenues for future research remain to enhance the effectiveness and robustness of these methods.

7.1 Adversarial attacks

One of the pressing challenges in fake news detection is the development of techniques to counter adversarial attacks. Adversaries can manipulate text to bypass detection models, making them vulnerable to subtle alterations. Future

Table 11 Performance of various models on the GossipCop dataset

Reference	Dataset	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score
Jarrahi and Safari 2023)	GossipCop	SLCNN	98.8	99.1	97.7	97.6%
TS, S.M., Sreeja, P. 2024)		Deep LSTM	94	–	–	–
Zhou et al. 2020)–		att–RNN	74.3	78.8	91.3	84.6
Galli et al. 2022)		Bi–LSTM	86	–	–	–
Palani et al. 2022)		CB–Fake	92	87	81	84%
Katarya et al. 2022)		MSVM	97.7	97.6	96	96.3%

Table 12 Performance of various models on the Pheme dataset

Reference	Dataset	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Jarrahi and Safari 2023)	Pheme	CNN	75.8	74	77	75.5
Katarya et al. 2022)		MSVM	88	–	–	–
Li et al. 2022)		LVDKF	74.6	86.5	80.1	84.9

Table 13 Performance of various models on the Weibo dataset

Reference	Dataset	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Wang et al. 2018)	Weibo	EANN	82.7	84.7	81.2	82.9
Khattar et al. 2019)		MVAE	82.4	85.4	76.9	82.9
–						
Katarya et al. 2022)		MSVM	94.7	–	–	–
Li et al. 2022)		LVDKF	96.1	92.4	94.2	93.9

Table 14 Performance measures for our deceptive-information detection system

Sr. no	Evaluation metric	Formula
1	Accuracy	$\frac{TN+TP}{TP+TN+FP+FN}$
2	Precision	$\frac{TP}{TP+FP}$
3	F1-score	$\frac{2*Precision*Recall}{Precision+Recall}$
4	Recall or true positive rate	$\frac{TP}{TP+FP}$
5	Error rate	$\frac{FN+FP}{TP+TN+FP+FN}$
6	Area under the curve	$\frac{1-FillOutRate+TruePositiveRate}{2}$
7	Miss rate	$\frac{FN}{TP+FN}$
8	Fall-our rate	$\frac{FP}{TN+FP}$
9	False discovery rate	$\frac{FP}{TP+FP}$
10	False omitted rate	$\frac{FN}{FN+TN}$
11	Specificity	$\frac{TN}{TN+FP}$

research should focus on creating models that are more resistant to such adversarial perturbations by integrating techniques from the field of adversarial machine learning.

7.2 Multi-modal analysis

Fake news is not limited to textual content; images, videos, and audio clips can also be manipulated to spread misinformation. Future research should explore ways to incorporate multi-modal analysis, which involves analyzing multiple forms of media to detect inconsistencies and anomalies that might indicate the presence of fake news.

7.3 Contextual understanding

Current fake news detection models often struggle with understanding the context in which a piece of information is presented. Enhancing models with contextual understanding, including the cultural, social, and historical aspects of a story, can improve their accuracy and reduce false positives.

7.4 Fine-grained labeling

Many fake news datasets currently use binary labels (fake/real). However, fake news exists on a spectrum, ranging from slight distortions to complete fabrications. Introducing finer-grained labels that capture the varying degrees of misinformation can aid in the development of more nuanced detection models.

7.5 Explainable AI

The interpretability of fake news detection models is crucial for building trust and understanding their decisions. Future research should focus on developing methods to make these models more explainable, allowing users to comprehend why a certain piece of content is flagged as fake.

7.6 Transfer learning

The effectiveness of fake news detection models can be limited by the availability of labeled data. Transfer learning techniques, where models trained on one domain are adapted to another with limited labeled data, could play a crucial role in addressing this issue.

7.7 Long-term dynamics

Fake news detection often focuses on the immediate detection of misinformation. How- ever, understanding the long-term dynamics of how fake news evolves and spreads can provide valuable insights into devising more effective strategies to counter its impact.

7.8 Ethical considerations

As fake news detection models become more powerful, ethical considerations surrounding privacy, bias, and unintended consequences become more crucial. Future research should address these concerns to ensure the responsible deployment of detection technologies (Tables 15 and 16).

7.9 Real-time detection

The speed at which fake news spreads demands real-time detection systems. Developing models that can analyze and flag potentially false information in real time is essential to mitigate the rapid dissemination of misinformation.

7.9.1 Recent advancements and models addressing challenges in fake news detection

Recent advancements in fake news detection have introduced several models to tackle the associated challenges. Hybrid

Table 15 Evaluating our survey in contrast to an established survey centered on social media platforms and varying utilized features

Reference	Year	Journal types	Approach	social networking platform	Varieties of features use by author
Zhou et al. 2020)	2020	Journal	Survey	Multiple	Content and context
Ahmed et al. 2018)	2018	Journal	Survey	Multiple	Content and context
Bharadwaj and Shao 2019)	2019	Journal	Survey	Multiple	Content and context
Reis et al. 2019)	2019	Journal	Other	Multiple	Content and context
Meinert et al. 2018)	2018	Conference	Survey	NA	Context and content
Shu et al. 2019)	2019	Journal	Survey	Multiple	Context, domain and content
Kaliyar et al. 2020)	2020	Journal	Other	NA	Context
Ren et al. 2016)	2016	Journal	Other	Twitter	Context
Tang et al. 2014)	2014	Journal	Other	Twitter	Context
Shahid et al. 2022)	2022	Journal	Survey	Twitter and facebook	Context and domain

Table 16 Advancements and models addressing challenges in fake news detection

Reference	Aspect	Details
Zhang and Ghorbani 2020)	Hybrid models	Hybrid models combining various techniques have shown promise, integrating deep learning with traditional machine learning approaches
Khalil et al. 2024)	CNN and LSTM combination	Models combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks leverage the strengths of both spatial feature extraction and temporal sequence processing
Essa et al. 2023)	BERT-based models	BERT-based models have been particularly effective due totheir contextual understanding, enhancing the detection accuracyof nuanced and context-dependent fake news
Mohapatra et al. 2022)	Graph-based approaches	Graph-based approaches have emerged to address the spreadand network influence of fake news, utilizing Graph NeuralNetworks (GNNs) to capture relational data between newsitems and their sources
Wang et al. 2021)	Multi-modal models	Multi-modal models incorporating text, images, and metadatahave also been developed to handle diverse forms of misinformation
Kumar and Taylor 2024)	Adversarial training methods	Adversarial training methods have been employed to makedetection models more robust against sophisticated fake newscrafted to evade detection

models (Zhang and Ghorbani 2020) combining various techniques have shown promise, integrating deep learning with traditional machine learning approaches. For instance, models combining Convolutional Neural Networks (CNNs) and Long Short- Term Memory (LSTM) networks (Khalil et al. 2024) leverage the strengths of both spatial feature extraction and temporal sequence processing. BERT-based models (Devlin et al. 1810; Essa et al. 2023) have been particularly effective due to their contextual understanding, enhancing the detection accuracy of nuanced and context-dependent fake news. Furthermore, graph-based approaches (Mohapatra et al. 2022; Monti et al. 1902) have emerged to address the spread and network influence of fake news, utilizing Graph Neural Networks (GNNs) to capture relational data between news items and their sources. Multi-modal models (Wang et al. 2021) incorporating text, images, and metadata have also been developed to handle diverse forms of misinformation. Additionally, adversarial training methods have been employed to make detection models more robust against sophisticated fake news crafted to evade detection. These models collectively address various challenges such as contextual understanding, relational data processing, and robustness against adversarial examples, providing a comprehensive approach to fake news detection on social media platforms.

Future work in the field of fake news detection could focus on several promising areas. One significant direction is the development and refinement of hybrid models that integrate deep learning and traditional machine learning approaches to further enhance accuracy and efficiency. This includes experimenting with different combinations of CNNs and LSTMs for better spatial and temporal feature extraction, as well as exploring advanced transformer models like BERT to improve contextual understanding of the LIAR dataset.

8 Conclusion

In the ever-evolving realm of information dissemination, the emergence of fake news has spurred a dynamic landscape of research and innovation in its detection. This survey paper delved into the recent trends and challenges within this critical domain. As evidenced by the strides made in recent years, the integration of advanced machine learning techniques, natural language processing, and network analysis has yielded promising results in identifying deceptive content. The collaborative efforts of researchers across disciplines have fostered a deeper understanding of the multifaceted nature of fake news, contributing to the refinement of detection models.

Nonetheless, the journey to effective fake news detection is riddled with complexities. Adversarial attacks persistently challenge the robustness of models, urging the need for adversarial training and enhanced security measures. The incorporation of multi-modal analysis, contextual nuances, and explainable AI offers a multi-dimensional approach to addressing the evolving tactics of misinformation.

Ethical considerations loom large, underscoring the importance of striking a balance between privacy, bias mitigation, and the responsible use of AI. The convergence of real-time detection capabilities represents an evolving frontier in the battle against fake news.

In summation, this survey illuminates the substantial progress made and the intricate challenges that lie ahead in the domain of fake news detection. As technology continues to shape the way information is disseminated, a concerted effort to navigate these challenges will pave the way for a more informed, trustworthy, and resilient information ecosystem.

Author Contribution All authors have read and approved the final manuscript and agree to be accountable for all aspects of the work, ensuring its accuracy and integrity.

Funding Not applicable.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest All authors declare that they have no conflicts of interest.

References

Ahmed H, Traore I, Saad S (2018) Detecting opinion spams and fake news using text classification. Secur Priv 1(1):9

Akdag SH, Cicekli NK (2024) Early detection of fake news on emerging topics through weak supervision. J Intell Inf Syst. https://doi.org/10.1007/s10844-024-00852-1

Athira A, Kumar SM, Chacko AM (2023) A systematic survey on explainable AI applied to fake news detection. Eng Appl Artif Intell 122:106087

Bharadwaj P, Shao Z (2019) Fake news detection with semantic features and text mining. Int J Natural Lang Comput (IJNLC) 8(17):22

Bra¸soveanu AM, Andonie R (2021) Integrating machine learning techniques in semantic fake news detection. Neural Process Lett 53(5):3055–3072. https://doi.org/10.1007/s11063-020-10365-x

Comito C, Caroprese L, Zumpano E (2023) Multimodal fake news detection on social media: a survey of deep learning techniques. Soc Netw Anal Min 13(1):101

Devlin J, Chang M-W, Lee K, Toutanova K (2018): Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805

Dong X, Victor U, Qian L (2020) Two-path deep semisupervised learning for timely fake news detection. IEEE Trans Comput Soc Syst 7(6):1386–1398

Elhadad MK, Li KF, Gebali F (2020) Detecting misleading information on covid- 19. Ieee Access 8:165201–165215

Essa E, Omar K, Alqahtani A (2023) Fake news detection based on a hybrid bert and lightgbm models. Complex Intell Syst 9(6):6581–6592

Faustini PHA, Covoes TF (2020) Fake news detection in multiple platforms and languages. Expert Syst Appl 158:113503

Fung Y, Thomas C, Reddy RG, Polisetty S, Ji H, Chang S-F, McKeown K, Bansal M, Sil A (2021) Infosurgeon: cross-media fine-grained information consistency checking for fake news detection. In: Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing pp 1683–1698

Galende BA, Hern´andez-Pen˜aloza G, Uribe S, Garcia FA (2022) Conspiracy or not? a deep learning approach to spot it on twitter. IEEE Access 10:38370–38378. https://doi.org/10.1109/ACCESS.2022.3165226

Galli A, Masciari E, Moscato V, Sperl´ı G (2022) A comprehensive benchmark for fake news detection. J Intell Inf Syst 59(1):237–261. https://doi.org/10.1007/s10844-021-00646-9

Hirlekar VV, Kumar A, (2020) Natural language processing based online fake news detection challenges–a detailed review. In: 2020 5th international conference on communication and electronics systems (ICCES), pp 748–754. IEEE

Huang Y-F, Chen P-H (2020) Fake news detection using an ensemble learning model based on self-adaptive harmony search algorithms. Expert Syst Appl 159:113584

Jain V, Kaliyar RK, Goswami A, Narang P, Sharma Y (2022) Aenet: an attention-enabled neural architecture for fake news detection using contextual features. Neural Comput Appl 34(1):771–782

Jarrahi A, Safari L (2023) Evaluating the effectiveness of publishers' features in fake news detection on social media. Multimed Tools Appl 82(2):2913–2939

Kaliyar RK, Goswami A, Narang P, Sinha S (2020) Fndnet–a deep convolutional neural network for fake news detection. Cogn Syst Res 61:32–44

Kaliyar RK, Goswami A, Narang P (2021a) Fakebert: fake news detection in social media with a bert-based deep learning approach. Multimed Tools Appl 80(8):11765–11788

Kaliyar RK, Goswami A, Narang P (2021b) Echofaked: improving fake news detection in social media with an efficient deep neural network. Neural Comput Appl 33:8597–8613

Karimi H, Roy P, Saba-Sadiya S, Tang J (2018) Multi-source multi-class fake news detection. In: proceedings of the 27th international conference on computational linguistics, pp 1546–1557

Katarya R, Dahiya D, Checker S et al (2022) Fake news detection system using featured-based optimized msvm classification. IEEE Access 10:113184–113199

Kausar S, Tahir B, Mehmood MA (2020) Prosoul: a framework to identify propaganda from online urdu content. IEEE Access 8:186039–186054

Khalil A, Jarrah M, Aldwairi M (2024) Hybrid neural network models for detecting fake news articles. Human-Centric Intell Syst 4(1):136–146

Khattar D, Goud JS, Gupta M, Varma V (2019): Mvae: Multimodal variational autoencoder for fake news detection. In: The World Wide Web Conference, pp 2915–2921

Kiruthika N, Rajagopalan T (2022) Dynamic light weight recommendation system for social networking analysis using a hybrid lstm-svm classifier algorithm. Opt Mem Neural Netw 31:59–75. https://doi.org/10.3103/S1060992X2201009X

Kumar A, Taylor JW (2024) Feature importance in the age of explainable AI: case study of detecting fake news and misinformation via a multi-modal framework. Eur J Oper Res 317(2):401–413

Li Q, Hu Q, Lu Y, Yang Y, Cheng J (2020) Multi-level word features based on cnn for fake news detection in cultural communication. Pers Ubiquit Comput 24:259–272

Li D, Guo H, Wang Z, Zheng Z (2021) Unsupervised fake news detection based on autoencoder. IEEE Access 9:29356–29365

Li K, Guo B, Liu J, Wang J, Ren H, Yi F, Yu Z (2022) Dynamic probabilistic graphical model for progressive fake news detection on social media platform. ACM Trans Intell Syst Technol (TIST) 13(5):1–24

Linmei H, Wei S, Zhao Z, Bin W (2022) Deep learning for fake news detection: a comprehensive survey. AI Open 3:133–155. https://doi.org/10.1016/j.aiopen.2022.09.001

Liu Y, Wu Y-FB (2020) Fned: a deep network for fake news early detection on social media. ACM Trans Inf Syst (TOIS) 38(3):1–33

Long Y, Lu Q, Xiang R, Li M, Huang C-R (2017) Fake news detection through multi-perspective speaker profiles. Proc Eighth Int Joint Confer Natural Lang Process 2:252–256

Ma J, Gao W, Wei Z, Lu Y, Wong K-F. (2015) Detect rumors using time series of social context information on microblogging websites. In: proceedings of the 24th ACM international on conference on information and knowledge management, pp 1751–1754

Malhotra P, Malik SK (2024) An efficient FTS-BERT based fake news detection using CKH_GANs classification technique. Multimed Tools Appl. https://doi.org/10.1007/s11042-024-19249-x

Meel P, Vishwakarma DK (2020) Fake news, rumor, information pollution in social media and web: a contemporary survey of state-of-the-arts, challenges and opportunities. Expert Syst Appl 153:112986

Meinert J, Mirbabaie M, Dungs S, Aker A (2018) Is it really fake?–towards an understanding of fake news in social media communication. In: social computing and social media. user experience and behavior: 10th international conference, SCSM 2018, held as part of HCI international 2018, Las Vegas, NV, USA, proceedings, part I 10, pp 484–497 Springer

Mohapatra A, Thota N, Prakasam P (2022) Fake news detection and classification using hybrid bilstm and self-attention model. Multimed Tools Appl 81(13):18503–18519

Monti F, Frasca F, Eynard D, Mannion D, Bronstein MM (2019) Fake news detection on social media using geometric deep learning. arXiv preprint arXiv:1902.06673

Mughaid A, Al-Z'ubi S, Al Arjan A, Al-Amrat R, Alajmi R, Zitar RA, Abualigah L (2022) An intelligent cybersecurity system for detecting fake news in social media websites. Soft Comput 26(12):5577–5591. https://doi.org/10.1007/s11042-022-12764-9

Nassif AB, Elnagar A, Elgendy O, Afadar Y (2022) Arabic fake news detection based on deep contextualized embedding models. Neural Comput Appl 34(18):16019–16032

Oliveira NR, Medeiros DS, Mattos DM (2020) A sensitive stylistic approach to identify fake news on social networking. IEEE Signal Process Lett 27:1250–1254

Orhan A (2023) Fake news detection on social media: the predictive role of university students' critical thinking dispositions and new media literacy. Smart Learn Environ 10(1):29

Oshikawa R, Qian J, Wang WY (2018) A survey on natural language processing for fake news detection. arXiv preprint arXiv:1811.00770

Ozbay FA, Alatas B (2021) Adaptive salp swarm optimization algorithms with inertia weights for novel fake news detection model in online social media. Multimed Tools Appl 80(26):34333–34357

P´erez-Rosas V, Kleinberg B, Lefevre A, Mihalcea R (2017) Automatic detection of fake news. arXiv preprint arXiv:1708.07104

Palani B, Elango S, Viswanathan K, V. (2022) CB-fake: a multimodal deep learning framework for automatic fake news detection using capsule neural network and bert. Multimed Tools Appl 81(4):5587–5620. https://doi.org/10.1007/s11042-021-11782-3

Rai N, Kumar D, Kaushik N, Raj C, Ali A (2022) Fake news classification using transformer based enhanced lstm and bert. Int J Cogn Comput Eng 3:98–105

Rajalaxmi RR, Narasimha Prasad LV, Janakiramaiah B, Pavankumar CS, Neelima N, Sathishkumar VE (2022) Optimizing hyperparameters and performance analysis of LSTM model in detecting fake news on social media. ACM Trans Asian Low-Res Lang Inf Process. https://doi.org/10.1145/3511897

Rani P, Jain V, Shokeen J, Balyan A (2022) Blockchain-based rumor detection approach for covid-19. J Ambient Intell Human Comput. https://doi.org/10.1007/s12652-022-03900-2

Rastogi S, Bansal D (2023) A review on fake news detection 3t's: typology, time of detection, taxonomies. Int J Inf Secur 22(1):177–212

Reddy H, Raj N, Gala M, Basava A (2020) Text-mining-based fake news detection using ensemble methods. Int J Autom Comput 17(2):210–221

Reis JC, Correia A, Murai F, Veloso A, Benevenuto F (2019) Supervised learning for fake news detection. IEEE Intell Syst 34(2):76–81

Ren Y, Wang R, Ji D (2016) A topic-enhanced word embedding for twitter sentiment classification. Inf Sci 369:188–198

Roy A, Basak K, Ekbal A, Bhattacharyya P (2018) A deep ensemble framework for fake news detection and classification. arXiv preprint arXiv:1811.04670

Sadeghi F, Bidgoly AJ, Amirkhani H (2022) Fake news detection on social media using a natural language inference approach. Multimed Tools Appl 81(23):33801–33821

Santia G, Williams J (2018) Buzzface: a news veracity dataset with facebook user commentary and egos. In: Proceedings of the International AAAI Conference on Web and Social Media, 12 pp 531–540

Shahid W, Li Y, Staples D, Amin G, Hakak S, Ghorbani A (2022) Are you a cyborg, bot or human?—a survey on detecting fake news spreaders. IEEE Access 10:27069–27083

Shan G, Zhao B, Clavin JR, Zhang H, Duan S (2021) Poligraph: intrusion- tolerant and distributed fake news detection system. IEEE Trans Inf Forensics Secur 17:28–41

Shishah W (2021) Fake news detection using bert model with joint learning. Arab J Sci Eng 46(9):9115–9127

Shu K, Sliva A, Wang S, Tang J, Liu H (2017) Fake news detection on social media: a data mining perspective. ACM SIGKDD Explor Newsl 19(1):22–36

Shu K, Mahudeswaran D, Liu H (2019) Fakenewstracker: a tool for fake news collection, detection, and visualization. Comput Math Organ Theory 25:60–71

Silva RM, Santos RL, Almeida TA, Pardo TA (2020) Towards automatically filtering fake news in portuguese. Expert Syst Appl 146:113199

Souza MC, Nogueira BM, Rossi RG, Marcacini RM, Dos Santos BN, Rezende SO (2022) A network-based positive and unlabeled learning approach for fake news detection. Mach Learn 111(10):3549–3592

Tang D, Wei F, Yang N, Zhou M, Liu T, Qin B. (2014) Learning sentiment-specific word embedding for twitter sentiment classification. In: proceedings of the 52nd annual meeting of the association for computational linguistics 1 pp 1555–1565

Truică C-O, Apostol E-S (2023) It's all in the embedding! fake news detection using document embeddings. Mathematics 11(3):508. https://doi.org/10.3390/math11030508

TS S M, Sreeja P (2024) Fake news detection on social media using adaptive optimization based deep learning approach. Multimed Tools Appl, 1–21

Umer M, Imtiaz Z, Ullah S, Mehmood A, Choi GS, On B-W (2020) Fake news stance detection using deep learning architecture (cnn-lstm). IEEE Access 8:156695–156706

Varghese SR, Juliet S, Athish N (2024) Social media text analysis for disaster management using distilbert model. In: 2024 international conference on science technology engineering and management (ICSTEM), 1–7. IEEE

Verma PK, Agrawal P, Amorim I, Prodan R (2021) Welfake: word embedding over linguistic features for fake news detection. IEEE Trans Comput Soc Syst 8(4):881–893

Wang WY (2017) "liar, liar pants on fire": a new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648

Wang Y, Wang L, Yang Y, Lian T (2021) Semseq4fd: integrating global semantic relationship and local sequential order to enhance text representation for fake news detection. Expert Syst Appl 166:114090

Wang Y, Ma F, Jin Z, Yuan Y, Xun G, Jha K, Su L, Gao J (2018) Eann: event adversarial neural networks for multi-modal fake news detection. In: Proceedings of the 24th acm sigkdd international conference on knowledge discovery and data mining pp 849–857

Ying L, Yu H, Wang J, Ji Y, Qian S (2021) Multi-level multi-modal cross- attention network for fake news detection. IEEE Access 9:132363–132373

Zhang X, Ghorbani AA (2020) An overview of online fake news: characterization, detection, and discussion. Inf Process Manage 57(2):102025

Zhao T, Wei M, Preston JS, Poon H (2023) Automatic calibration and error correction for large language models via pareto optimal self-supervision. arXiv preprint arXiv:2306.16564

Zhou Xinyi, Jindi Wu, Zafarani Reza (2020) SAFE: Similarity-aware multi-modal fake news detection. In: Lauw Hady W, Wong Raymond Chi-Wing, Ntoulas Alexandros, Lim Ee-Peng, Ng See-Kiong, Pan SinnoJialin (eds) Advances in knowledge discovery and data mining: 24th Pacific-Asia Conference, PAKDD 2020, Singapore, May 11–14, 2020, Proceedings, Part II. Springer International Publishing, Cham, pp 354–367. https://doi.org/10.1007/978-3-030-47436-2_27

Zubiaga A, Liakata M, Procter R 2017 Exploiting context for rumour detection in social media. In: social informatics: 9th international conference, socinfo 2017, Oxford, UK, proceedings, pp 109–123 Springer

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