International Journal of Computer Sciences and Engineering
Vol.12, Issue.7, pp.48-52, July 2024
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

ISSN: 2347-2693 (E)

Review Article

Advancements and Challenges in Fake News Detection using Machine Learning: A Comprehensive Review

Avnis Kumar[1*], Chetan Agrawal[2], Pooja Meena[3]

[1,2,3]Dept. of CSE, Radharaman Institute of Technology and Science, Bhopal (M.P.), India

**Corresponding Author: aksingh91096@gmail.com*

Received: 20/May/2024; **Accepted**: 22/Jun/2024; **Published**: 31/Jul/2024. **DOI**: https://doi.org/10.26438/ijcse/v12i7.4852

Abstract: The rapid proliferation of fake news across digital platforms has emerged as a challenging task, undermining public discourse, and compromising public trust in media. Initially, the detection efforts focused on textual features using traditional machine learning algorithms, which, despite their effectiveness, were limited by the manual and time-consuming process of feature extraction. The advent of deep learning heralded a significant shift, with Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) offering enhanced capabilities in capturing the nuanced interplay of textual elements. Parallelly, the examination of visual features through multimodal methods demonstrated the importance of incorporating images and videos, further refined by Graph Convolutional Networks (GCNs) and attention mechanisms for superior accuracy. However, challenges persist in integrating and fully utilizing multimodal information, particularly in addressing the limitations of deep versus shallow feature analysis and the adaptability of models across diverse scenarios. This paper synthesizes the methodologies, findings, and critical evaluations of these approaches, highlighting the advancements and identifying areas for future research in the detection of fake news.

Keywords: Fake News Detection, Textual Feature Extraction, Visual Feature Analysis, Multimodal Analysis, Machine Learning Algorithms, Deep Learning.

1. Introduction

The utilization of social networks has led to a massive influx of information, serving as the primary channel of communication globally [1]. While social networks facilitate the rapid dissemination of news, distinguishing between real and fake news poses a significant challenge. The spread of fake news on social media platforms can have far-reaching and detrimental effects, making the timely detection and containment of such misinformation a critical necessity [2]. Fake news, widely recognized for its impact on public opinion and societal events, is characterized by its intent to mislead through the dissemination of false or misleading information. Definitions of fake news vary, with some emphasizing its intentionality and verifiability as false, and others highlighting its mimicry of legitimate news media formats without adherence to journalistic processes or intent [3]. Common characteristics of fake news include the echo chamber effect, where biased information is amplified within insular groups, and the intention to deceive, often driven by political, ideological, or financial motives. Additionally, the role of malicious accounts, including social bots, trolls, and cyborg users, is critical in the creation and spread of fake news, exploiting social media platforms to manipulate public opinion or provoke emotional responses [4]. Fake news often comprises nonfactual statements and aims to generate confusion and distrust among the public, making it difficult to discern reality from falsehood. Various related concepts, such as disinformation, rumours, and misinformation, share overlapping characteristics with fake news but differ in intentions and the nature of the statements made [5]. Fake news detection is the process of identifying whether a news article is genuine or false, leveraging both content-based features, typical of traditional media, and social context-based features, more relevant to social media platforms. This detection involves a formal approach that uses social interactions among users to predict the authenticity of a news article. Defined as a binary classification task, fake news detection aims to classify a news article as either fake (1) or not fake (0) based on a prediction function F, which evaluates the content and context of the news. However, beyond simple binary classification, fake news detection can also be approached as a multiclassification task, where news articles are classified into multiple categories based on a broader set of labels. This approach acknowledges the complexity of misinformation which require advance detection and classification methods. The paper proposes a comprehensive definition of fake news as containing nonfactual statements, intended to mislead the public and create echo chamber effects. This paper is used to define the role of machine learning for fake news detection and its related complexity of addressing misinformation in the digital age and highlights the importance of developing effective detection and mitigation strategies [6].

Therefore, fake news refers to misinformation that is intentionally created to deceive readers, often mimicking the style of legitimate news outlets. Its characteristics include the spread within echo chambers—closed networks where similar viewpoints are amplified while excluding others—and its creation by malicious actors like social bots and trolls [7]-[15]. The detection and management of fake news is crucial due to its potential to manipulate public opinion, influence political processes, and cause societal harm. With millions globally relying on social media for news, the impact of fake news is substantial, influencing elections, public health responses, and other critical areas. Understanding and combating fake news is vital for maintaining informed societies and the integrity of democratic processes.

Motivated by this, the purpose of this survey paper is to provide a comprehensive analysis of the methodologies, challenges, and advancements in detecting fake news. The objectives include:

- To outline the definitions and characteristics of fake news.
- To review the technological and psychological factors contributing to its spread.
- To evaluate current strategies and tools used for fake news detection.
- To identify gaps in research that could be addressed to improve detection techniques.

Rest part of the paper is organized as: Section 2 presents the background study of the detection model. Section 3 presents the literature review. Section 4 presents the current challenges and future scope. Finally conclusion is presented in section 5.

2. Background of Fake News Detection Models

Fake news has been a part of human civilization since its inception, but modern technologies and changes in the global media landscape have significantly accelerated its spread. The consequences of fake news are profound, affecting social, political, and economic environments. It shapes our perceptions and decisions, leading to critical misjudgements when based on distorted or fabricated information found online. The primary impacts of fake news include damage to individuals who may face harassment or threats on social media, health misinformation which poses severe risks given the rise in people seeking health information online, financial repercussions as seen in manipulated stock prices or tarnished business reputations, and democratic implications, notably illustrated by the influence of fake news in the American presidential election. These areas highlight the urgent need to curb the spread of fake news to mitigate its real-life detrimental effects [16].

Fact-checking is a vital method for detecting fake news, traditionally relying on either expert-based or crowd-sourced approaches. Expert-based fact-checking utilizes highly credible, specialized fact-checkers and is known for its accuracy but struggles with scalability due to high costs and management challenges. Many expert-based fact-checking websites, such as PolitiFact and HoaxSlayer, offer in-depth analysis and have contributed to creating datasets for fake news research. On the other hand, crowd-sourced fact-checking taps into the collective intelligence of a large population, offering broader scalability but facing issues with credibility and accuracy due to political biases and inconsistent results. Platforms like Amazon Mechanical Turk and emerging sites like Fiskkit illustrate this approach, providing a space where users can rate and tag articles, enhancing the understanding of patterns in fake versus genuine content. As major social media platforms recognize the importance of combating misinformation, more crowd-sourced tools are expected to develop, supporting the detection and analysis of fake news across different media [17].

Automatic fact-checking has emerged as a necessary response to the scalability limitations of manual methods, given the vast amount of information continuously generated, particularly on social media. This approach leverages advancements in Information Retrieval (IR), Natural Language Processing (NLP), Machine Learning (ML), and network/graph theory to automate the verification process. The process starts with extracting facts from diverse sources to construct a knowledge base (KB), which involves collecting, cleaning, and resolving inconsistencies such as redundancy, invalidity, conflicts, unreliability, and incompleteness of data. These tasks are supported by sophisticated models and methods that predict links and validate facts against established knowledge graphs (KGs). Fact-checking then compares extracted data (structured as triples of Subject, Predicate, Object) with facts in the KG to determine their authenticity. This comparison relies on assumptions like closed-world (non-existing triples are false), open-world (non-existing could be true or false), and local closed-world (authenticity is determined based on the presence of related triples). This systematic approach facilitates the efficient evaluation of news authenticity, harnessing large-scale existing KBs and cutting-edge analytical techniques to meet the demands of rapid information dissemination [17][18].

3. Methodology

This literature review adopts a systematic methodology to address research questions, focusing on the role of machine learning in fake or false news detection. To conduct this review, research papers were selected from various databases to ensure a comprehensive exploration of the topic. The inclusion and exclusion criteria were used for the selection process. Papers not written in English, inaccessible in full, or unrelated to machine learning and fake news detection were excluded. Conversely, papers that were accessible, in English, and relevant to the topic were included. The quality of the selected papers was assessed based on their contributions to the field of machine learning for detecting fake or false news, ensuring that only high-quality papers were discussed and cited in the review. This methodological approach helps in effectively answering the research questions posed, providing a structured and focused analysis of the existing literature.

4. Literature Review

The detection of fake news initially relied heavily on analyzing textual features within articles, using methods focused on statistical or semantic levels of text feature extraction, such as the number of paragraphs, lexical percentage, symbols, writing style, and language style. Traditional machine learning algorithms applied these extracted features to detect fake news, but this manual feature extraction process proved to be challenging, time-consuming, and often unable to fully leverage the text content. To address these limitations, deep learning methods, known for their potent representation learning capabilities, were adopted. Models using recurrent neural networks (RNNs) and convolutional neural networks (CNNs) have been developed to capture the nuances of text features over time and to examine deeper interactions between key features, respectively, demonstrating significant improvements over traditional machine learning outcomes in fake news detection. Alongside textual analysis, the examination of visual features in news content has gained prominence with the advent of multimodal news formats. Studies have explored the role of images and their types in disinformation detection, yet the initial methods showed significant limitations due to their simplistic learning modes. Innovations in this area include CNN-based models designed to identify image patterns and frequency-domain characteristics, augmented by semantic detection and attention mechanisms for integrating visual and frequency domain patterns. While the inclusion of visual features has enhanced the accuracy of fake news detection, challenges remain regarding the adaptability and effectiveness of these models across various scenarios. Rachna et al. [1] explored the use of LSTM combined with Random Forest (RF) on a dataset named "ban fake news," capturing essential hidden clues in tweet texts but achieving a relatively low accuracy rate of approximately 63%. This study highlighted the need for further exploration into user profiles and textual features. On the other hand, Kao et al. [2] applied multi-view attention networks to the Twitter15 and Twitter16 datasets, significantly improving accuracy to around 93%, though the model struggled with analyzing replies and comments on tweets. Verma et al. [3] introduced the WELFake method on BuzzFeed News, utilizing Word Embedding over Linguistic Features to achieve a high accuracy rate of 96%. However, their approach did not consider knowledge graphs and user credibility. Similarly, Xu et al. [4] employed Latent Dirichlet Allocation on BuzzFeed News to distinguish between the domain reputations and topics of news, observing real and fake news similarities but not performing textual analysis. Sansonetti et al. [5] detected unreliable users on social media platforms like PolitiFact.com using Deep Neural Networks, achieving nearly 93% accuracy despite a high learning loss rate. Shahbazi and Byun [6] leveraged Natural Language Processing (NLP) and Blockchain technology for fake news detection on BuzzFeed News, achieving a Mean Absolute Percentage Error (MAPE) of approximately 1.87. However, their method was hindered by high latency (~1500ms) and a high Root Mean Square Error (RMSE) (~318), indicating areas for potential improvement. Tuan and Minh [7] developed a method for efficiently learning and fusing multimodal features from posts. Similarly, Song et al. [8] introduced the KMGCN framework, which utilizes Graph Convolutional Networks to fuse textual, visual, and knowledge information for fake news detection. Khattar et al. [9] combined text and image modalities using multimodal variant encoders. Giachanou et al. [10] enhanced fake news detection by merging semantic representations from image-article title similarities with features learned by VGG-16 and BERT. Wang et al. [11] designed an event adversarial neural network (EANN) focusing on learning feature representations of text and images. Despite these advancements, many studies have primarily considered the spatial domain information of images, overlooking the potential insights from frequency domain analysis. This gap becomes particularly relevant as CNN-based generated image models become more sophisticated, making it challenging to distinguish between real and synthesized images. To address these challenges, Giachanou et al. [12] proposed SceneFND, integrating text with contextual scenes and visual representations for improved detection performance. Gôlo et al. [13] introduced MVAE-FakeNews, a multimodal approach that enhances fake news detection by integrating text embeddings, topic, and linguistic information through Optimal Composite Learning (OCL). Tested on real-world datasets in Portuguese and English, MVAE-FakeNews demonstrated superior performance in terms of F1-Score and AUC-ROC across three datasets, outperforming fourteen other methods, and showed competitive results in three additional datasets. Remarkably, it achieved comparable or better outcomes with only 3% of labeled fake news data. The study also introduced Multimodal LIME for OCL, aiming to discern how each modality contributes to fake news classification. Yin et al. [14] developed the Multi-modal Co-Attention Capsules Network (MCCN), structured around feature extraction, fusion, and classification layers. This network effectively builds features from user profiles, multimodal source news, and comments. It employs a dual Cross-Modal Co-Attentional mechanism for integrating multi-modal interactions between text and images, and a Hierarchical Co-Attention for fusing user information with source news content and comments. The classification is executed using a capsules network, demonstrating excellent performance across three major datasets when compared to other baselines. Yang et al. [15] proposed the multimodal relationship-aware attention network (MRAN), which undertakes a three-step process for fake news detection. Initially, it employs a multi-level encoding network and VGG19 to extract hierarchical semantic and visual features, respectively. Then, these representations are processed through a relationship-aware attention network to generate high-order fusion features by assessing intra-modal and cross-modal similarities. Lastly, these features are utilized by a fake news detector for identification. This structure underscores the importance of analyzing and integrating multiple modalities and their interrelations for more accurate fake news detection. Despite these significant strides in multimodal fake news detection, a common limitation is the focus on deep feature correlations across modalities, often at the expense of exploiting shallow feature information. This oversight suggests a potential avenue for further enhancing the performance of fake news detection systems by incorporating a more holistic analysis of both deep and shallow multimodal features. Below Table 1 presents the critical review analysis of the presented literature.

Table 1. Critical Review Analysis

Ref	Method	Key Findings	Strengths	Limitations
Rachna et al. [1]	LSTM + RF	Achieved ~63% accuracy.	Captured important hidden clues in tweet texts.	Low accuracy; lacks user profile and textual feature exploration.
Kao et al. [2]	Multi-view attention networks	Achieved ~93% accuracy.	Significantly improved accuracy; advanced attention mechanisms.	Struggles with analyzing replies and comments.
Verma et al. [3]	WELFake	Achieved 96% accuracy.	Utilized Word Embedding over Linguistic Features for high accuracy.	Did not consider knowledge graphs and user credibility.
Xu et al. [4]	Latent Dirichlet Allocation	Observed domain reputations and news topics.	Distinguished between real and fake news similarities.	Lacked comprehensive textual analysis.
Sansonetti et al. [5]	Deep Neural Networks	Achieved nearly 93% accuracy.	Detected unreliable users on social media.	High learning loss rate.
Shahbazi and Byun [6]	NLP + Blockchain	Achieved MAPE of ~1.87.	Adopted innovative methods for detection.	High latency (~1500ms) and high RMSE (~318).
Tuan and Minh [7]	Multimodal feature fusion	Enhanced multimodal feature learning.	Efficiently fused multimodal features.	Complex
Song et al. [8]	KMGCN framework	Utilized GCNs for fusing textual, visual, and knowledge information.	Advanced integration of multimodal data.	Complex
Khattar et al. [9]	Multimodal variant encoders	Combined text and image modalities.	Simple fusion of modal information.	Complex
Giachanou et al. [10]	VGG-16 + BERT	Merged semantic representations with visual and text features.	Enhanced detection with image-article title similarities.	Low Efficiency
Giachanou et al. [12]	SceneFND	Integrated text with contextual scenes and visual representations.	Improved performance with contextual scenes.	Complex
Gôlo et al. [13]	MVAE-FakeNews	Demonstrated superior F1-Score and AUC-ROC.	Integrated text embeddings, topic, and linguistic information.	Focused mainly on deep feature correlations.
Yin et al. [14]	MCCN	Achieved excellent performance on large-scale datasets.	Utilized a dual Co-Attentional mechanism for feature fusion.	Complex
Yang et al. [15]	MRAN	Identified fake news through multimodal relationship-aware attention.	Extracted hierarchical semantic and visual features.	Focused on deep feature correlations; potential oversight of shallow features.

The research presented on fake news detection have used different approaches and methodologies but it also shows some significant gaps and problems.

- Many models, like the LSTM + RF by Rachna et al. [1], do not incorporate user profile data, which can be crucial for understanding the credibility and influence of information sources.
- As noted in Kao et al. [2] have focused on high-performing models but they struggle with analyzing interactions such as replies and comments, which are significant in the spread of information on social media.
- High complexity in models like the KMGCN framework by Song et al. [8] and MCCN by Yin et al. [14] does not always translate to significantly better performance or practical applicability due to increased computational demands and implementation difficulties.
- Some models achieve high accuracy but suffer from low efficiency, as seen in the VGG-16 + BERT model by Giachanou et al. [10]. This raises questions about the feasibility of deploying such models in real-time systems where quick processing is crucial.

5. Current Challenges and Future Scope

Based on the literature provided, the current challenges and future scope in the domain of fake news detection are multifaceted, reflecting both the complexity of the problem and the rapid evolution of technologies and methodologies aimed at addressing it. Some models, such as the one proposed by Rachna et al. (2021), achieve relatively low accuracy (around 63%) when dealing with intricate details embedded within the text, indicating a need for more sophisticated models that can capture and analyze nuanced information more effectively. The inability to analyze replies or comments, as seen in the model by Kao et al. (2021), suggests a significant gap in assessing the broader context of news dissemination and reception on social media platforms. The high latency (~1500ms) and computational requirements noted in the approach by Shahbazi and Byun (2020) pose significant challenges for real-time fake news detection and mitigation, especially on large-scale social media platforms. The absence of detailed textual and contextual analysis in some studies indicates a need for more in-depth content scrutiny, beyond simple topic or domain reputation assessments.

Future research could focus on incorporating multimodal data (text, images, videos) to provide a more holistic and accurate detection of fake news, considering the increasing prevalence of multimedia content in news dissemination. By analyzing user profiles, engagement patterns, and network structures, models can better understand the propagation mechanisms of fake news and identify key influencers in spreading misinformation. The adoption of more sophisticated NLP and AI methodologies, such as transformer-based models, could significantly improve the understanding and detection of nuanced and sophisticated fake news narratives. Developing methods that reduce latency and computational demands can facilitate real-time detection of fake news, a crucial factor in preventing the widespread dissemination of misinformation.

6. Conclusion

The fake news detection has undergone significant transformations, moving from reliance on textual analysis to embracing the complexities of multimodal data. The evolution from traditional machine learning to deep learning has unlocked new potentials in accurately identifying fake news, leveraging the rich representations of text and visual data. Recent studies have each contributed to this body of knowledge, showcasing the effectiveness of various models in different contexts. The integration of textual, visual, and sometimes even knowledge-based information through advanced models like MVAE-FakeNews, MCCN, and MRAN represents the cutting edge of current research, offering promising results but also revealing the necessity for further innovation. Particularly, the challenge of effectively combining deep and shallow features, alongside ensuring the adaptability of detection models to new and evolving forms of fake news, remains a critical area for future exploration. This review underscores the significant strides made in fake news detection as multifaceted approach to research, one that is quick and responsive to the ever-changing digital landscape.

References

[1] Jain, Rachna & Jain, Deepak & Dharana, & Sharma, Nitika. Fake News Classification: A Quantitative Research Description. ACM Transactions on Asian and Low-Resource Language Information Processing, **2022.** 21. 1-17. 10.1145/3447650.

[2] S. Ni, J. Li and H. -Y. Kao, "MVAN: Multi-View Attention Networks for Fake News Detection on Social Media," in IEEE Access, Vol.**9**, pp.**106907-106917**, **2021**. doi: 10.1109/ACCESS.2021.3100245.

[3] P. K. Verma, P. Agrawal, I. Amorim and R. Prodan, "WELFake: Word Embedding Over Linguistic Features for Fake News Detection," in IEEE Transactions on Computational Social Systems, Vol.**8**, No.**4**, pp.**881-893**, **2021**. doi: 10.1109/TCSS.2021.3068519.

[4] K. Xu, F. Wang, H. Wang and B. Yang, "Detecting fake news over online social media via domain reputations and content understanding," in Tsinghua Science and Technology, Vol.**25**, No.**1**, pp.**20-27**, **2020**. doi: 10.26599/TST.2018.9010139.

[5] G. Sansonetti, F. Gasparetti, G. D'aniello and A. Micarelli, "Unreliable Users Detection in Social Media: Deep Learning Techniques for Automatic Detection," in IEEE Access, Vol.**8**, pp.**213154-213167**, **2020**. doi: 10.1109/ACCESS.2020.3040604.

[6] Z. Shahbazi and Y. -C. Byun, "Fake Media Detection Based on Natural Language Processing and Blockchain Approaches," in IEEE Access, Vol.**9**, pp.**128442-128453**, **2021.** doi: 10.1109/ACCESS.2021.3112607.

[7] N. M. Duc Tuan and P. Quang Nhat Minh, "Multimodal Fusion with BERT and Attention Mechanism for Fake News Detection," 2021 RIVF International Conference on Computing and Communication Technologies (RIVF), Hanoi, Vietnam, pp.**1-6, 2021.** doi: 10.1109/RIVF51545.2021.9642125.

[8] Song, Chenguang, et al. "Knowledge augmented transformer for adversarial multidomain multiclassification multimodal fake news detection." Neurocomputing 462, pp.**88-100, 2021.**

[9] Khattar, Dhruv, et al. "Mvae: Multimodal variational autoencoder for fake news detection." The world wide web conference. **2019.**

[10] Giachanou, Anastasia, Guobiao Zhang, and Paolo Rosso. "Multimodal fake news detection with textual, visual and semantic information." Text, Speech, and Dialogue: 23rd International Conference, TSD 2020, Brno, Czech Republic, September, pp.**8–11, 2020**. Proceedings 23. Springer International Publishing, 2020.

[11] Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., et al., Eann: Event adversarial neural networks for multi-modal fake news detection. In Proceedings of the 24th Acm sigkdd international conference on knowledge discovery & data mining, pp.**849–857, 2018.**

[12] Zhang, Guobiao, Anastasia Giachanou, and Paolo Rosso. "SceneFND: Multimodal fake news detection by modelling scene context information." Journal of Information Science (2022): 01655515221087683.

[13] Gôlo, Marcos Paulo Silva, et al. "One-class learning for fake news detection through multimodal variational autoencoders." Engineering Applications of Artificial Intelligence 122, **2023**: 106088.

[14] Yin, Chunyan, and Yongheng Chen. "Multi-Modal Co-Attention Capsule Network for Fake News Detection." Optical Memory and Neural Networks 33.1, pp.**13-27, 2024.**

[15] Yang, Hongyu, et al. "MRAN: Multimodal relationship-aware attention network for fake news detection." Computer Standards & Interfaces 89, 103822, **2024.**

[16] Mridha, Muhammad F., et al. "A comprehensive review on fake news detection with deep learning." IEEE access 9, pp.**156151-156170, 2021.**

[17] Zhou, Xinyi, and Reza Zafarani. "A survey of fake news: Fundamental theories, detection methods, and opportunities." ACM Computing Surveys (CSUR) 53.5, pp.**1-40, 2020.**

[18] Hangloo, Sakshini, and Bhavna Arora. "Fake News Detection Tools and Methods--A Review." arXiv preprint arXiv:2112.11185, **2021**.