International Journal of Advanced Engineering Research and Science (IJAERS)
Peer-Reviewed Journal
ISSN: 2349-6495(P) | 2456-1908(O)
Vol-11, Issue-7; Jul, 2024
Journal Home Page Available: https://ijaers.com/
Article DOI: https://dx.doi.org/10.22161/ijaers.117.11

A Comprehensive Review of Multi-Modal Fusion Strategies in Fake News Detection

Dhramveer Moga[1], Dr. Vijay Kumar[2]

[1]Research Scholar, Jayoti Vidyapeeth Women's University, Jaipur, India
dvs.moga@gmail.com
[2]Professor, Jayoti Vidyapeeth Women's University, India
drvijay@jvwu.ac.in

Received: 30 May 2024,

Receive in revised form: 20 Jun 2024,

Accepted: 25 Jul 2024,

Available online: 30 Jul 2024

***Keywords*— Fake News Detection, Multi-Modal Fusion, Deep Learning, Misinformation, Early Fusion, Late Fusion, Hybrid Models, CNN, Transformers, Explainable AI, Adversarial Robustness, Fact-Checking.**

***Abstract*—** *Fake news spreads across various digital platforms around the world, posing a significant challenge, requiring far more sensitive and advanced detection techniques beyond those used in traditional text-based methodology. This inspired multi-modal fusion strategies that merge textual, visual, and contextual information as a modern strategy to enhance detection of fake news. This review provides a comprehensive analysis of state-of-the-art multi-modal fusion techniques, categorizing them into early, late, and hybrid fusion strategies. We look into the trend of deep learning architectures such as CNN, RNN, transformers, and attention in leveraging multi-modal information for better classification accuracy, as well as widely used datasets, evaluation benchmarks, and real-world applications in social media fact-checking and misinformation mitigation. Recent advancements notwithstanding, challenges include data scarcity, computational complexity, bias, and interpretability. We highlight open research directions, focusing on explainable AI, adversarial robustness, and scalable cross-platform detection methods. This review is intended to be a foundational resource for researchers and practitioners who are trying to develop robust multi-modal frameworks for combating misinformation in the digital age.*

I. INTRODUCTION

The rapid expansion of digital communication platforms has facilitated the swift dissemination of information across the globe. Although this has made news and knowledge more accessible, it has also led to the widespread spreading of misinformation and disinformation, which is generally referred to as "fake news [1]. Fake news spreads across various digital platforms around the world, posing a significant challenge, requiring far more sensitive and advanced detection techniques beyond those used in traditional text-based methodology [1]. This inspired multi-modal fusion strategies that merge textual, visual, and contextual information as a modern strategy to enhance detection of fake news. This review provides an exhaustive analysis of state-of-the-art multi-modal fusion techniques categorized into early, late, and hybrid fusion strategies. We examine the trend of deep learning architectures like CNN, RNN, transformers, and attention in using multi-modal information to improve classification accuracy, as well as widely used datasets, evaluation benchmarks, and real-world applications in social media fact-checking and misinformation mitigation [2]. Recent advancements notwithstanding, some of the challenges include data scarcity, computational complexity, bias, and interpretability. We identify open research directions in the area, emphasizing explainable AI, adversarial robustness, and scalable cross-platform detection methods. This survey is meant to be a resource for researchers and practitioners who try to develop robust multi-modal

frameworks to combat the problem of misinformation in the digital age [2].

A. Scope of the Paper

Misinformation across the web has been one of the hottest trends lately and has brought about the limitations that traditional text-based fake news detection methods have imposed. In reality, as the multimedia elements increase their role in deceptive content in the form of manipulated images, videos, or synthetic audio, there is a need for something more effective and comprehensive in its detection strategy. This paper focuses on multi-modal fake news detection [3] where various data sources such as text, images, and contextual data are aggregated to improve accuracy and reliability in detecting misinformation. The multi-modal learning techniques within the fake news detection system provide a more integrated analysis of the content, indicating inconsistencies across multiple modalities which may not come out when strictly analyzing a source.This review summarizes diverse topics encompassing multi-modal fake news detection, providing profound multi-modal fusion strategy analysis such as early fusion and late where differences in combined approaches determine a more effective blend for better detecting multi-modal based. It further speaks about advanced architectures of deep learning, such as CNNs, RNNs, transformers, and attention-based mechanisms [3], all of which heavily contributed to multi-modal misinformation detection success. Then, benchmarking datasets and assessment metrics are critically reviewed, considering the strengths and weaknesses of each set, highlighting the need for more diverse, large-sized datasets. Beyond theoretical analysis, this paper is also a case study on some real-world applications, including fact-checking for social media, automated detection of misinformation in journalism, and minimizing the spread of false information within political and health-related domains.

While this review aims to provide a comprehensive discussion on AI-driven multi-modal fake news detection, it does not focus on purely text-based approaches or theoretical misinformation studies from a sociological perspective. Instead, it emphasizes the role of deep learning and AI-driven techniques in integrating different modalities for enhanced misinformation detection. This scope ensures that the paper remains relevant to researchers and practitioners working on the development of advanced fake news detection systems.

B. Objectives of the Paper

The overall objective of this review is to give a structured and detailed view of multi-modal fake news detection [4], emphasizing methodologies, developments, and challenges within the area. The paper begins by categorizing and analyzing current multi-modal fusion strategies in light of the methods applied, that is, early fusion, feature-level integration, late fusion, decision-level combination, and a mix of both, as contributory to increased accuracy in detection. The review appropriately compares the different methods and sheds significant light on the respective merits and demerits of various types of fusion strategies.

Another main goal is discussing deep learning architectures that power multi-modal misinformation detection. Several improved neural networks, especially CNNs for analyzing images, transformers for processing text, and attention mechanisms for cross-modal interactions, have significantly improved the performance of models for detecting fake news. This review discusses how all these architectures are used in various multi-modal settings and discusses trends in terms of GANs [4] for detecting deepfake-based misinformation.

Furthermore, this paper aims to evaluate the datasets and benchmarks commonly used in multi-modal fake news detection. While several datasets have been proposed for this task, many suffer from issues such as data imbalance, limited modality coverage, or lack of real-world diversity. By reviewing existing datasets and their associated evaluation metrics, this paper identifies key gaps and suggests areas for improvement in dataset creation and benchmarking.

Besides the technical aspects, it also looks forward to the usage of such review in real case studies and beyond, as most multi-modal systems are implemented to be deployed for practical usage, especially social media platforms, news verification agencies, and automated fact-checking systems, all being increasingly reliant on AI-driven methodologies to fight false information. Here, by shedding light on a few use cases, the work provides a real-world view towards the efficiency of multi-modal detection techniques [5].

Even with all these new improvements, challenges remain, including issues related to open research questions that remain unanswered within this multi-modal fake news detection framework. Examples include issues on data scarcity, computational complexity, adversarial robustness, and model interpretability, among others. These aspects will be explained in depth with proposed solutions, followed by further considerations related to the ethics involved in developing these responsible AI solutions: issues regarding algorithmic bias and transparency on AI-driven models detecting misinformation [5].

Finally, this paper aims to highlight future research directions, emphasizing the need for scalable and explainable AI models for fake news detection. With the increasing sophistication of misinformation tactics—such as AI-generated fake news articles and deepfake videos [6]—future research must focus on developing more robust, interpretable, and generalizable detection frameworks. Areas such as cross-lingual and cross-platform fake news detection, adversarial resistance, and real-time misinformation tracking present promising avenues for further exploration.

These review objectives will then be used by this review in providing a starting point for all researchers, policymakers, and technologists interested in advancing multi-modal fake news detection. It brings together a coherent overview of available methodologies while noting key challenges as well as new opportunities for innovative solutions in tackling misinformation in this digital age.

II. BACKGROUND AND FUNDAMENTALS

A. Concept of Fake News

Fake news can be defined as the intentional dissemination of false or misleading information presented as actual news to influence the public, to proclaim specific ideas or ideologies, or for financial or political gains. It is generally differentiated from misinformation since the latter might be an honest mistake or misunderstanding, but the former is made with a malicious intent to mislead. It can be partial fabricated stories, altered factual content, misattributed images and videos, or misleading narratives that distort real-world events. There has been an amplified effect brought about by the digital age because these respective inventions allow this to spread at a fast pace [6].

Such a content often relies on sensationalism and emotional appeal as one of its defining characteristics, which means false stories are often created to invoke very powerful feelings like fear, anger, or outrage with the intention to share it without checking it for the users [6]. Fake news also often lacks credible sources or distorts them by selectively quoting data out of context. As most elements of multimedia will be manipulated, doctored images, edited videos, or AI-generated deepfakes, this makes it much more challenging to discern such contents. Fake news exploits confirmation bias by the fact that people easily believe and share news more readily if it conforms to what they already think is true rather than what they actually know.

From public confidence in media to distorted democratic process and worsening social unrest, fake news knocks all these ones into the ground. Unchecked and unrestrained diffusion of misinformation in fields like politics, public health, and finance might have terrible consequences in actual life, which include vaccine hesitancy, stock market manipulation, and elections interference. This has led to the detection of false information being a serious area of research that is gaining a lot of interest towards the application of artificial intelligence and machine learning in the war against misinformation [7].

B. Challenges in Fake News Detection

Detecting fake news is inherently complex due to the evolving nature of misinformation tactics. One of the primary challenges lies in the dynamic and adaptive strategies employed by misinformation creators [8]. The arms race between fact-checkers and spammers has to be acknowledged since the development of detection models enhances adversarial techniques that will outsmart the latter. Deepfakes and other AI-generated synthetic news articles present another problem to be overcome with detection as this content can convincingly imitate real news items [8].

Other prominent challenges include linguistic ambiguity and contextual deception. Most hoaxes combine facts and fiction in ways that even the text-based models have difficulty distinguishing the true from the false stories. Satirical and sarcastic articles also complicate matters since it becomes hard for the algorithms to catch the intended meanings and undertones of the communication. Secondly, misleading information draws on legitimate news events through some contextual distortions, hence further complicating how automated systems detect inconsistencies [9].

A major deficiency of existing state-of-the-art research on fake news detection is that there is no large-scale high-quality labeled dataset. Although a lot of datasets are available, most of them tend to be skewed to some extent, favoring specific domains, such as political misinformation, or lack important multi-modal elements for complete analysis. Misinformation also arises very frequently in new forms and languages, so generalizability to perform well on diverse contexts pose a large challenge in building the model [9].

Scalability and real-time detection pose huge challenges. Fake news spreads extremely quickly, sometimes achieving viral levels before mechanisms of verification can help balance it. Other challenges are models that scale hugely in dealing with multi-modal content in a non-computationally expensive manner, and there's the challenge of adversarial attacks, slight word changes or distortion in the images, misleading metadata, and many more. These require much more resilient and adaptive approaches, which can easily beat the detection systems in place.

Given these challenges, there is an increasing move toward multi-modal fake news detection in which various types of data including text, images, videos, and metadata may be used together to provide much more accurate and robust approaches toward the detection of misinformation [9].

C. Role of Multi-Modal Data: Text, Image, Video, Audio, Metadata

Since it's no longer just plain texts alone, but lately, people always consider such multiple modal data such as images and videos, or even sound recordings for that matter, for better credibility with engagement, the detection method of fake news has also reached a multi-modality where everything is put up together in order to carefully consider the pattern against misinformation. Given that there has been analysis done in terms of how the modalities interact with one another, then detectors should easily pick the matches which single-modal methods don't normally see [10].

Though based on text alone is where fake news' base of foundation does rest, forms of fake news flow mostly via articles, social media, comments. Most traditional methods that detect fake news employ NLP techniques, such as: sentiment analysis, stylistic feature extraction, and transformer-based models like BERT and GPT [10]. These methods evaluate linguistic patterns, writing styles, and semantic coherence regarding the deception. Sometimes, however, text-based analysis alone is not enough, especially when pictures or even videos are used to support the falsity [11].

Pictures and videos are also part of the false reports. Manipulated images or misattributed photographs characterize most fake news articles. Sometimes, even videos depicting no one or nothing at all are reutilized. For instance, a completely irrelevant picture of a natural disaster might be assigned to a recently reported incident to inspire fear. So far, CNNs and ViTs were commonly used for the image and video analysis involved in detecting inconsistencies, manipulated parts, or visual deepfakes. These enable detection models to look at pixel-level patterns that would be indicative of manipulation signatures [11].

Deepfake audio and synthetic speech add yet another layer of complexity to the detection of fake news. Voices that mimic public figures are extremely difficult to tell whether they are real or fake audio. Audio-based disinformation, therefore, poses a dangerous situation in political and financial contexts where false statements ascribed to some prominent individuals are used to create a certain view of things among the public. Techniques using speech recognition and deep learning to analyze audio include spectrogram-based CNNs and recurrent neural networks [12].

Metadata, including timestamps of publication, user engagement metrics, and source credibility scores, also gives crucial clues in the detection of fake news. Misinformation is mostly due to coming from low-credibility sources, bot accounts, or any coordinated disinformation campaign. The detection of such misinformation is carried out by making use of metadata analysis. Adding metadata along with textual, visual, and auditory information may improve the accuracy of the multi-modal models designed for fake news detection [12].

In general, multi-modal data integration supports system robustness in detecting fake news because they would cross-reference several sources of information. On the whole, multi-modal detection approaches that include text, images, videos, audio, and metadata are better at characterizing inconsistencies as well as manipulative content and harmful narratives as compared to the classic single-modality approaches. Multi-modal anti-digital deception efforts will require much more prominence against the advancing methods of misinformation.

Multimodal deep learning models for detecting fake news

The increasing incidence of fake news in multi-modal text, images, videos, and audio has naturally led to using deep learning techniques for the detection of misinformation. Several models designed for the task include CNNs, RNNs, Transformers, Attention Mechanisms, and GANs, to improve the accuracy in detecting multi-modal misinformation [13].

CNNs are mainly used to process images and videos, find inconsistencies like manipulated images, deepfakes, and unnatural visual elements. When combined with word embedding techniques, they also are used in text analysis to find deceptive writing patterns and misleading visual content [13].

RNNs and Transformers are effective for sequential data processing, so they are really important for fake news detection, especially in the textual and speech-based formats. RNN captures long-range dependencies in text. Transformers, BERT, GPT, or other models using self-attention mechanisms enhance the contextual understanding process.Vision-and-language Transformers even further enhance the multi-modal detection by analyzing the relationship between the textual claims and the visual evidence [14].

Combination of several modalities is done by attention mechanisms that focus on the most important features of an object and filter out misleading information. Cross-

modal attention happens specifically to align text with an image or video thus can help detect inconsistencies and verify authentic claims [15].

Generative Adversarial Networks (GANs) represent both a problem and a solution for fake news detection. Although GANs have commonly been popularly applied to generate extremely realistic deepfake videos and synthetic news articles, they also represent powerful tools for detecting manipulated media [16]. Adversarial training enables GAN-based models to pick subtle differences in the deepfake images and videos and helps improve the detection. Models for detection would continue to grow with regard to countering strategies of increasingly sophisticated misinformation being introduced by GANs. Therefore, deep learning models have made significantly better advances into multi-modal fake news detection based on the analysis of images via CNNs and text or speech processing through the use of RNNs or Transformers. Using attention mechanisms will further advance this to allow more multi-modal fusion. In that regard, it has become much more significant because GANs can further promote adversarial learning [16].

III. METHODOLOGY

The review paper deals with a systemic review and summation of state-of-the-art developments in the deep learning-based techniques for detecting multi-modal fake news. A structured literature review is part of the methodology by selecting relevant studies, strict criteria for screening them, and analysis of the research contributions that would have the highest impact. These are the steps that outline the methodology used within this review:

A. Relevant studies

Relevant studies published in the renowned, established digital libraries and academic databases like IEEE Xplore, Springer, Elsevier, Wiley Online Library, ACM Digital Library, and Google Scholar have been used as a source. In addition to that, leading artificial intelligence and NLP conferences such as the Conference on Neural Information Processing Systems (NeurIPS), International Conference on Machine Learning (ICML), the Association for Computational Linguistics (ACL) and the Conference on Computer Vision and Pattern Recognition (CVPR) are adapted through peer-reviewed paper selection in order to take care of the recent advances that deep learning-based approaches have made in detecting fake news.

B. Selection Criteria

Studies were chosen based on inclusion criteria of relevance and quality up to the review.

- Time Period: Only those studies which were conducted within the last ten years between 2014-2024 are included, thus providing the latest updates; however, pre-study phases conducted before the given period have also been considered in this paper especially when such studies yield core theoretical or methodological breakthroughs.
- The relevance to fake news detection is that only such studies have been taken which focused only on the detection of fake news by using multi-modal approaches. This review did not allow studies dealing with general misinformation without much focus on multi-modal fusion.
- Orientation of Deep Learning: Most of the studies in the analyses were based on deep learning techniques, which include CNNs, RNNs, Transformers, Attention Mechanisms, and GANs. Machine learning has not been used except when it was successfully integrated with deep learning models, such as SVMs or Decision Trees.
- Multi-Modal Data Usage: Those studies, which only dealt with the incorporation of multiple modalities, namely, text, image, video, audio, and metadata are taken into account. Single-modal fake news detection researches which involve only models like text-only and image-only unless any data concerning multimodal integration are out of the list because of these criteria.

C. Keyword Strategy

A keyword-based strategic search was used to obtain relevant literature. The following keywords were used applying Boolean operators AND, OR so as to minimize the results.

- Multi-modal fake news detection
- Deep learning for misinformation detection
- AI-based media forensics
- Convolutional Neural Networks (CNN) in the process of fake news detection
- Transformers models and analysis of misinformation GANs and deepfakes
- Multimodal Fusion Strategies in AI
- Fake news detection via text and image
- Attention in multi-modal learning

D. Selection Procedure

The search was conducted at three levels:

- Primary Search: An initial reconnaissance was conducted on the databases that had been identified using the keywords chosen. This returned about 60 peer-reviewed articles relevant to the search. The

relevant citing articles were also traced through citation tracking.

- Shortlisting: After screening, according to pre-decided criteria, only those papers collected, which possessed their titles and abstracts; as a result of this filtering stage, the subsequent 30 top-quality paper entries were finalized toward studies involving relevance to deep learning and multi-modal fake news detection.
- Final Selection and Review: For each of the 20 papers, an in-depth analysis has been performed. Contributions of the articles to deep learning models, novel multi-modal fusion approaches, or even a real-world application in detecting misinformation will be analyzed in detail. All these papers represented the central themes of this review.

Proceeding with the systematic approach, the review presents a general overview of the current progresses, challenges, and future directions of research related to multi-modal fake news detection using deep learning techniques.

Table 1. Selection Process of Literature Review

Stage	Number of Papers	Description
Initial Collection	100	Papers identified through database searches and citation tracking using defined keywords.
Shortlisting	25	Papers reviewed for relevance based on inclusion criteria (multi-modal fake news detection, deep learning models, and AI techniques).
Final Review	15	In-depth review of selected papers focusing on CNNs, transformers, attention mechanisms, and multi-modal fusion strategies in fake news detection.

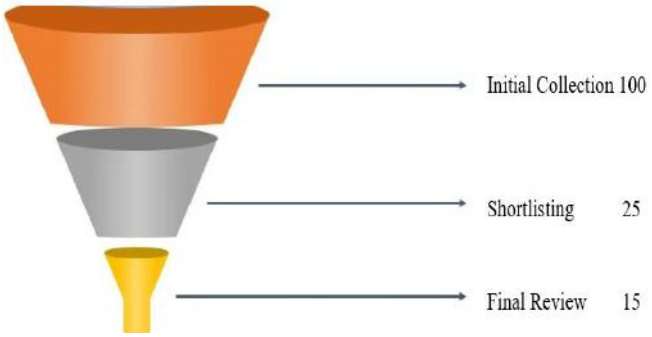

Fig.1. Funnel Diagram for Literature Review

IV. LITERATURE REVIEW

The final selected papers for literature review are as follows,

Utilizing the Fakeddit dataset, **Luqman et al. (2024)** [17] proposed an ensemble learning-based approach for identifying multimodal fake news and studied various risks that arise with rapid diffusion of fake news in modern technological society by means of social media. They proceeded to pre-process textual content with NLP to extract sentiment and to create text and images embeddings with Visual-BERT. Finally, using a 10-fold evaluation approach, an ensemble model is learned that yields much higher improvements on accuracy, precision, recall, F1-score, MCC, and OR against state-of-the-art approaches.

Kumar, A. & Taylor, J. W. (2024) [18]. Explainable multimodal fake news detection: Feature engineering for improved explanations. Their approach extracts textual as well as visual features, in the form of article titles and content, in addition to their images, along with combining their multimodal features using feature engineering to represent their behavior in propogating the fake news; the approach also outperformed single-modality methods on benchmarks without relying upon feature optimization techniques.

In **2024, Fu, L., et al.**[19] proposed TD-MMC, which is the first multimodal system for fake news detection. This includes text dominance, text-image complementarity, and cross-modal inconsistencies. A better text representation using social network data with the filtering of irrelevant information based on unidirectional cross-modal attention improved the model. Better results have been found based on public English and Chinese datasets.

Ayetiran, E. F., &Özgöbek, Ö. 2024 [20]. Interrelation of Fake News, Hate Speech, and Offensive Language: Towards Multimodal Approaches Ayetiran, E. F., &Özgöbek, Ö. 2024). By presenting the idea of image-text representation along with the traditional models, deep neural network using the inter-modal attention is proposed to enhance prediction. Their model outperforms prior works across three standard datasets, demonstrating the effectiveness of multimodal integration in detecting harmful content.

A recent paper by **Padalko, H. et al. (2024) [21]** reviews deep learning techniques to determine fake news. It particularly looks at the Bi-LSTM and attention-based Bi-LSTM models. In their attention model, the authors include attention mechanisms to evaluate the importance of various elements in input. The comparative studies indicated a 97.66% accuracy rate for their attention-based Bi-LSTM. There is a dependency on data, which may result in

overfitting; however, the advanced technique presented here deals with misinformation.

Dong, Y., et al. (2024, March) [22] proposed the NSLM model that can introduce three types of deception patterns, including image manipulation, cross-modal inconsistency, and image repurposing. These models apply amortized variational inference and pseudo-Siamese networks for capturing these deception patterns. Evaluations on real-world datasets show that NSLM achieved detection accuracy superior to others and gives insightful explanations of fake news fabrication.

Dev, D. G., et al. (2024) [23] identified challenges in the detection of fake news, particularly limited resources, such as the availability of datasets and computational tools. They promoted feature abstraction and vectorization by using Python's scikit-learn module for better numerical representation of text data. Their research focused on methods of feature selection to improve model accuracy and efficiency in identifying fake news.

Shan, F. et al., 2024,[24] "Proposal of a Multimodal Model Based on Similarity Reasoning and Adversarial Networks for Fake News Detection", utilizes BERT and Text-CNN to extract textual features and VGG-19 to extract visual features. The accuracy of fake news detection has increased to 86% on datasets from Twitter, but the paper admits that a more holistic model would be more complete by exploring other modalities besides text and images.

Hashmi, E., et al. (2024) [25] highlights the arising issue of misinformation via social media, and the need for sufficient fake news detection techniques in media-intensive content. The hybrid model uses FastText word embeddings together with machine learning/deep learning techniques such as CNNs and LSTM, with optimized hyperparameters to avert overfitting. The hybrid model had very high accuracy on three public datasets, with F1-scores of 0.99, 0.97, and 0.99. Other transformer models, including BERT, XLNet, and RoBERTa, performed well, and it was explained using AI.

Specifically, authors **Madani et al. [26]** experimented in 2024 on a wide variety of datasets that illustrate a good number of challenge setups related to the problem of un-informativeness in detection of fake news. They have used concepts of two-stage models with the intention to learn structural features and used curriculum strategy approaches by employing the k-nearest neighbor algorithm, which resulted in better accuracy for the detection of fake news as compared to various benchmarks.

Truică, C. O., et al. (2024) [27] report that apart from these requirements, the context-aware detection is also required. The architecture of DANES is based on the deep neural network ensemble; it takes social and textual contexts into consideration. Preliminary experiments on real-world datasets BuzzFace, Twitter15, and Twitter16 have revealed that DANES outperforms existing solutions, even trained on limited amounts of data.

To address problems such as biased sample distribution, **Zhou, Q., & Cai, T. (2024) [28]** presented a multi-scale recurrent convolutional neural network of the Electra-based form for the detection of fake news. Using gated residual connections and self-attention mechanisms, the model could possibly identify malicious news.

Kishwar, A., & Zafar, A. (2023) [29] proposed a study on the development of fake news in Pakistan. The authors were well prepared and have tested varied AI techniques in addition to the GloVe and BERT embeddings, CNN and LSTM and in addition machine learning methods like Naive Bayes and SVM also. Of all these the best performance of the model came through LSTM with GloVe whereas from the comparison another insight was derived also through the human judgment as well.

Mallick, C., et al. (2023) [30] presented a prototype model to estimate the news reliability by incorporating user feedback applying a cooperative deep learning approach. The proposed model may attain an accuracy of about 98% in false news detection as against many other language-processing models.

Alarfaj, F. K., & Khan, J. A. (2023) [31] have proposed work on the machine learning and deep learning approach towards fake news detection. It evaluates a number of models over the "Fake News" dataset and reveals that the deep learning model, like LSTM and CNN-LSTM, has superior capabilities in comparison with other traditional algorithms when it comes to the fake news classification domain.

Table 1. Literature Review Findings

Author Name (Year)	Main Concept	Findings	Limitations
Luqman et al. (2024)	Ensemble learning-based detection using multimodal data	Achieved significant improvements in accuracy, precision, recall, and F1-score using NLP and Visual-BERT embeddings	Limited to the Fakeddit dataset, may not generalize to all forms of fake news
Kumar &Taylor	Explainable AI for fake news detection with feature	Outperformed single-modality models by using multimodal feature	Lacks feature optimization

(2024)	engineering	extraction techniques	techniques
Fu et al. (2024)	TD-MMC framework integrating text-image complementarity	Improved multimodal fake news detection using text dominance and social network data	Limited to English and Chinese datasets
Ayetiran&Özgöbek (2024)	Multimodal detection of fake news, hate speech, and offensive language	Deep learning with inter-modal attention outperformed prior models	Challenges in effectively modeling complementarity between modalities
Padalko et al. (2024)	Deep learning with Bi-LSTM and attention-based Bi-LSTM	Achieved 97.66% accuracy in fake news classification	Dependency on data, risk of overfitting
Dong et al. (2024)	NSLM model with deceptive pattern detection	Identified image manipulation, cross-modal inconsistency, and repurposing patterns in fake news	Requires weak supervision for accurate results
Dev et al. (2024)	Feature abstraction and vectorization for fake news detection	Used scikit-learn for better text representation, improving accuracy	Computational resource constraints
Shan et al. (2024)	Multimodal model with similarity reasoning and adversarial networks	Achieved 86% accuracy in Twitter dataset using BERT and Text-CNN	Limited to text and image modalities
Hashmi et al. (2024)	Hybrid deep learning model with CNN and LSTM	High accuracy (F1-score of 0.99) using FastText embeddings	Model may need further generalization across datasets
Madani et al. (2024)	Two-stage models with curriculum learning for fake news detection	Improved accuracy using KNN-based feature learning	Challenge in handling uninformative data
Truică et al. (2024)	Context-aware deep neural ensemble (DANES)	Outperformed existing models even with limited data	Requires more real-world testing for diverse social contexts
Zhou & Cai (2024)	Multi-scale recurrent CNN with Electra-based model	Gated residual connections and self-attention mechanisms improved detection	Biased sample distribution affects generalization
Kishwar & Zafar (2023)	Fake news detection in Pakistan using AI techniques	LSTM with GloVe embeddings performed best	Human judgment still plays a crucial role
Mallick et al. (2023)	Cooperative deep learning approach for news reliability	Achieved 98% accuracy using user feedback	Dependent on user participation for effectiveness
Alarfaj& Khan (2023)	Machine learning and deep learning models for fake news detection	CNN-LSTM outperformed traditional classifiers	Requires further evaluation on different datasets

Research gaps Discussion

Despite the excellent work done on multimodal fake news detection, there are still many research gaps. Most of the strategies, for example, Luqman et al. (2024) and Kumar & Taylor (2024), aimed at improving the accuracy of the model through ensemble learning and feature engineering, do not generalize well to any dataset. In addition, though TD-MMC (Fu et al., 2024) as well as DANES (Truică et al., 2024) extract textual, visual, and social context, there still exist formidable challenges in effectively capturing the cross-modal inconsistency and context-aware misinformation. Dependency on particular datasets, as in

Hashmi et al. (2024) and Madani et al. (2024), creates problems such as model robustness and bias and, therefore, limits real-world applicability. Additionally, models that are based on attention, like attention-based Bi-LSTM (Padalko et al., 2024) and Electra-based CNN (Zhou & Cai, 2024), suffer from overfitting, which necessitates generalization techniques. Although deep learning techniques, such as CNN-LSTM (Alarfaj& Khan, 2023) and hybrid models (Hashmi et al., 2024), have achieved high accuracy, they consume a lot of computational resources, which makes real-time deployment difficult. Moreover, very few studies focus on real dynamics of misinformation propagation, behaviors of users, and adversarial robustness. This study shows existence of gaps in explainability and interpretability. Future work should be more holistic covering the incorporation of real-world social network interactions, multi-modal fusion techniques, and explainable AI methods to enhance the reliability and transparency of models for fake news detection.

V. CONCLUSION

Multi-modal fake news detection has been recognized as a critical area of research in the current spread of misinformation across digital platforms. Therefore, this review establishes the importance and significance of deep learning approaches, such as CNNs, RNNs, transformers, attention mechanisms, and GANs, in modeling and integration of diversified modalities, including text images, videos, audio, and metadata. Despite their outstanding detection performance improvements, such methods face a multitude of problems related to scarcity in data, robustness to adversaries, and dynamism of false information. It is very imperative that further studies address improvements on multi-modal fusion strategies, improved explainability on AI-based models, and detection systems on the fly, where misinformation can change in style according to increasingly cunning misinformation approaches. Deep learning methodologies may advance, combined with interdisciplinary approaches to fight the issue of fake news, so it can effectively help in developing a more trustworthy and reliable digital information ecosystem.

REFERENCES

[1] Qu, Z., Meng, Y., Muhammad, G., & Tiwari, P. (2024). QMFND: A quantum multimodal fusion-based fake news detection model for social media. *Information Fusion*, *104*, 102172.

[2] Zaheer, H., & Bashir, M. (2024). Detecting fake news for COVID-19 using deep learning: a review. *Multimedia Tools and Applications*, 1-34.

[3] Chang, Q., Li, X., & Duan, Z. (2024). Graph global attention network with memory: A deep learning approach for fake news detection. *Neural Networks*, *172*, 106115.

[4] Farokhian, M., Rafe, V., & Veisi, H. (2024). Fake news detection using dual BERT deep neural networks. *Multimedia Tools and Applications*, *83*(15), 43831-43848.

[5] Mallik, A., & Kumar, S. (2024). Word2Vec and LSTM based deep learning technique for context-free fake news detection. *Multimedia Tools and Applications*, *83*(1), 919-940.

[6] Luvembe, A. M., Li, W., Li, S., Liu, F., & Wu, X. (2024). CAF-ODNN: Complementary attention fusion with optimized deep neural network for multimodal fake news detection. *Information Processing & Management*, *61*(3), 103653.

[7] Singh, B., & Sharma, D. K. (2022). Predicting image credibility in fake news over social media using multi-modal approach. *Neural Computing and Applications*, *34*(24), 21503-21517.

[8] Uppada, S. K., & Patel, P. (2023). An image and text-based multimodal model for detecting fake news in OSN's. *Journal of Intelligent Information Systems*, *61*(2), 367-393.

[9] Hu, L., Wei, S., Zhao, Z., & Wu, B. (2022). Deep learning for fake news detection: A comprehensive survey. *AI open*, *3*, 133-155.

[10] Hangloo, S., & Arora, B. (2022). Combating multimodal fake news on social media: methods, datasets, and future perspective. *Multimedia systems*, *28*(6), 2391-2422.

[11] Jing, J., Wu, H., Sun, J., Fang, X., & Zhang, H. (2023). Multimodal fake news detection via progressive fusion networks. *Information processing & management*, *60*(1), 103120.

[12] Alghamdi, J., Lin, Y., & Luo, S. (2022). A comparative study of machine learning and deep learning techniques for fake news detection. *Information*, *13*(12), 576.

[13] Chen, M. Y., Lai, Y. W., & Lian, J. W. (2023). Using deep learning models to detect fake news about COVID-19. *ACM Transactions on Internet Technology*, *23*(2), 1-23.

[14] Segura-Bedmar, I., & Alonso-Bartolome, S. (2022). Multimodal fake news detection. *Information*, *13*(6), 284.

[15] Zhang, Q., Guo, Z., Zhu, Y., Vijayakumar, P., Castiglione, A., & Gupta, B. B. (2023). A deep learning-based fast fake news detection model for cyber-physical social services. *Pattern Recognition Letters*, *168*, 31-38.

[16] Sedik, A., Abohany, A. A., Sallam, K. M., Munasinghe, K., & Medhat, T. (2022). Deep fake news detection system based on concatenated and recurrent modalities. Expert Systems with Applications, 208, 117953.

[17] Luqman, M., Faheem, M., Ramay, W. Y., Saeed, M. K., & Ahmad, M. B. (2024). Utilizing ensemble learning for detecting multi-modal fake news. *IEEE Access*.

[18] Kumar, A., & Taylor, J. W. (2024). Feature importance in the age of explainable AI: Case study of detecting fake news

& misinformation via a multi-modal framework. *European Journal of Operational Research*, *317*(2), 401-413.

[19] Fu, L., Peng, H., Ma, C., & Liu, Y. (2024). Fake News Detection Based on Text-Modal Dominance and Fusing Multiple Multi-Model Clues. *Computers, Materials & Continua*, *78*(3).

[20] Ayetiran, E. F., &Özgöbek, Ö. (2024). An inter-modal attention-based deep learning framework using unified modality for multimodal fake news, hate speech and offensive language detection. *Information Systems*, *123*, 102378.

[21] Padalko, H., Chomko, V., &Chumachenko, D. (2024). A novel approach to fake news classification using LSTM-based deep learning models. *Frontiers in big Data*, *6*, 1320800.

[22] Dong, Y., He, D., Wang, X., Jin, Y., Ge, M., Yang, C., & Jin, D. (2024, March). Unveiling Implicit Deceptive Patterns in Multi-Modal Fake News via Neuro-Symbolic Reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 8, pp. 8354-8362).

[23] Dev, D. G., Bhatnagar, V., Bhati, B. S., Gupta, M., &Nanthaamornphong, A. (2024). LSTMCNN: A hybrid machine learning model to unmask fake news. *Heliyon*, *10*(3).

[24] Shan, F., Sun, H., & Wang, M. (2024). Multimodal Social Media Fake News Detection Based on Similarity Inference and Adversarial Networks. *Computers, Materials & Continua*, *79*(1).

[25] Hashmi, E., Yayilgan, S. Y., Yamin, M. M., Ali, S., &Abomhara, M. (2024). Advancing fake news detection: hybrid deep learning with fasttext and explainable AI. *IEEE Access*.

[26] Madani, M., Motameni, H., & Roshani, R. (2024). Fake news detection using feature extraction, natural language processing, curriculum learning, and deep learning. *International Journal of Information Technology & Decision Making*, *23*(03), 1063-1098.

[27] Truică, C. O., Apostol, E. S., & Karras, P. (2024). DANES: Deep neural network ensemble architecture for social and textual context-aware fake news detection. *Knowledge-Based Systems*, *294*, 111715.

[28] Zhou, Q., & Cai, T. (2024). Adaptive gate residual connection and multi-scale RCNN for fake news detection. *Machine Learning with Applications*, 100612.

[29] Kishwar, A., & Zafar, A. (2023). Fake news detection on Pakistani news using machine learning and deep learning. *Expert Systems with Applications*, *211*, 118558.

[30] Mallick, C., Mishra, S., & Senapati, M. R. (2023). A cooperative deep learning model for fake news detection in online social networks. *Journal of Ambient Intelligence and Humanized Computing*, *14*(4), 4451-4460.

[31] Alarfaj, F. K., & Khan, J. A. (2023). Deep Dive into Fake News Detection: Feature-Centric Classification with Ensemble and Deep Learning Methods. *Algorithms*, *16*(11), 507.