Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 5, No.3, May : 2025

A COMPREHENSIVE REVIEW ON COMPUTATIONAL TECHNIQUES FOR MULTI-MODAL FAKE NEWS DETECTION FOR ENGLISH LANGUAGE

Yash Dwivedi, JD College of Engineering and Management, Nagpur
Sandesh Mate, JD College of Engineering and Management, Nagpur
Sonal Dharmik, JD College of Engineering and Management, Nagpur
Prathamesh Bangade JD College of Engineering and Management, Nagpur

ABSTRACT

The proliferation of fake news across online platforms poses a significant threat to societal well-being and informed decision-making [1]. While early detection efforts primarily focused on textual content, the increasing prevalence of multi-modal misinformation, which strategically combines text and images, necessitates more sophisticated approaches [2]. This research investigates the landscape of multi-modal fake news detection, aiming to provide a comprehensive understanding of current methodologies, challenges, and future directions. We analyze various techniques that leverage the complementary information from textual and visual modalities, including feature extraction, cross-modal interaction modeling, and fusion strategies. Furthermore, we delve into the inherent complexities of this task, such as semantic gaps between modalities, the potential for subtle manipulation of visual content, and the scarcity of large-scale, high-quality multi-model datasets. By examining recent advancements and open challenges, this paper underscores the importance of developing robust and interpretable multi-modal fake news detection systems to mitigate the spread of online deception. The insights presented contribute to a more nuanced understanding of this critical area and pave the way for future research to enhance the accuracy and reliability of fake news detection in the multi-modal information ecosystem.

I. Introduction

The digital age has fundamentally transformed information production, sharing, and consumption [3]. This transformation, driven by the internet and social media, has increased news accessibility and facilitated its rapid global spread, often amplified by user re-sharing. However, this ease of dissemination has also fostered a significant and growing problem: the generation and propagation of fake news and rumors [4]. A notable portion of news on platforms like Twitter, Facebook, Instagram, and WhatsApp is estimated to be false, and this unchecked spread of misinformation can have detrimental societal effects, impacting public perception, decision-making, social interactions, and media views. Consequently, the development of effective fake news detection methods has become a critical concern for researchers.

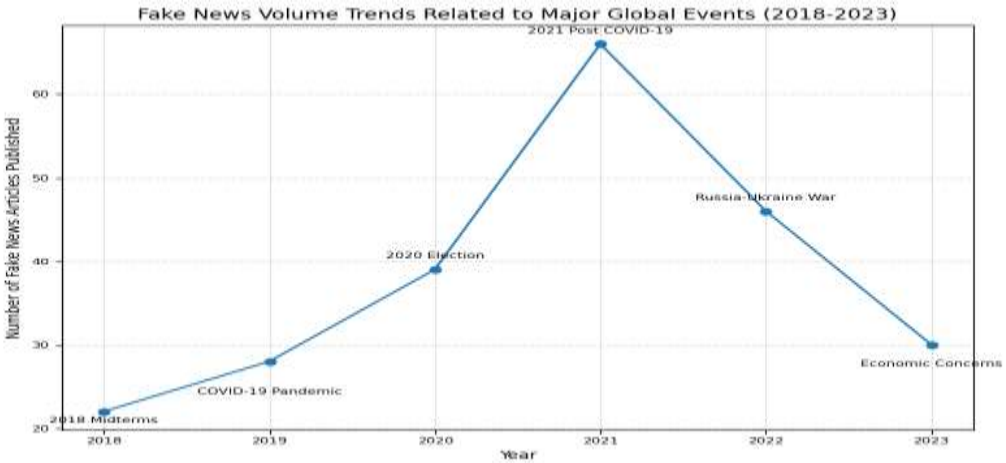

Figure 1: Fake News Volume Trends Related to Major Global Events (2018-2023)

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 5, No.3, May : 2025

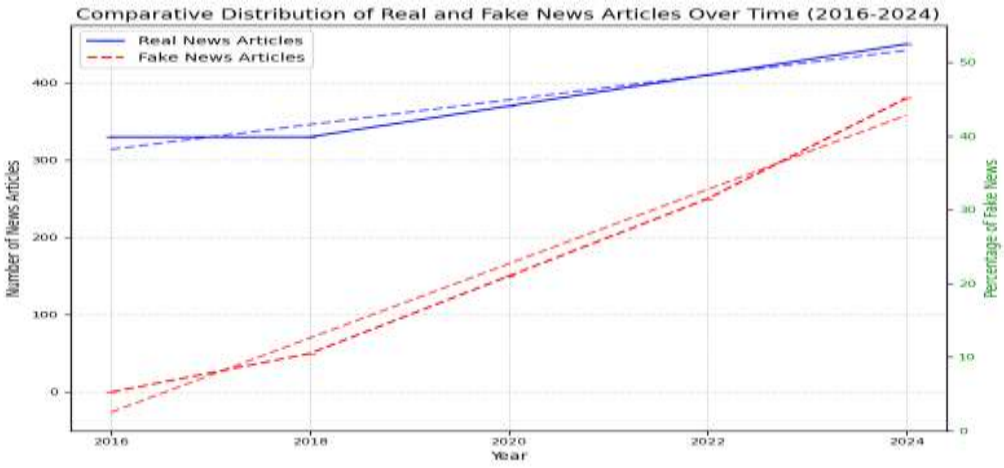

Figure 2: Comparative Distribution of Real and Fake News Articles Over Time (2016-2024)

Addressing fake news presents several multifaceted challenges:

- Inherent Complexity: Fake news can manifest in various forms and styles, making it difficult to create universal detection rules [5].
- Multiplicity of Sources and Intentions: It can originate from diverse sources with varying malicious intentions, including political agendas, business motives, and even cyberbullying [6].
- Multi-modality: The increasing use of combined text and images necessitates techniques that can understand the interplay between different information formats [7].
- Difficulties in Fact-Checking and Annotation: Manually verifying the truthfulness of a large volume of information is time-consuming and costly, posing a significant bottleneck for dataset creation [8].

Fake news can also be generated by machines, sometimes through the alteration of human-written news, a process that can result in machine-generated fake news overlooking key features that humans typically consider. Fake news detection is commonly approached as a binary classification task, categorizing articles as either true (¡span class=”math-inline”¿y 1¡/span¿) or fake (¡span class=”math-inline”¿y 0¡/span¿), though some approaches use multi-class categorization to reflect the stance of a news item (e.g., ’Disagree’, ’Agree’, ’Unrelated’, ’Discuss’). Initiatives like the First False News Challenge (FNC-1), involving numerous industry and academic groups, have focused on creating AI and machine learning-based automated detection systems, with one objective being to determine the stance of media production concerning specific titles.

Researchers have explored diverse fake news detection approaches, including data mining, machine learning (ML), computational intelligence, deep learning (DL), and social network analysis [9]. These methods often involve extracting and analyzing features from news content and its context. Analysis can be content-based, focusing on the text itself, or context-based and propagation-based, examining how news spreads across social media. Multimodal approaches are also employed, integrating features from different modalities like text and images, which is increasingly important given the growth of multi-modal content and the benefits of cross-modal learning. The following section will delve into specific content-based detection techniques.

II. Literature

2.1 Content-Based Detection

Textual analysis forms the foundation of early fake news detection efforts. Classical Natural Language Processing (NLP) techniques such as n-gram extraction, Term Frequency Inverse Document Frequency (TF-IDF), syntactic parsing, and sentiment analysis formed the early pipelines for text-based fake news detection. For instance, analyzing the frequency of specific word combinations (n-grams) can reveal stylistic patterns indicative of fabricated content.

Recent breakthroughs in pre-trained language models such as Word2Vec [11], GloVe [12], and

BERT [13] have allowed models to learn contextual semantics, outperforming rule-based systems. These embedding-based methods have demonstrated the ability to capture both syntactic and semantic nuances essential to discerning fake from real news. Word2Vec, for example, represents words as vectors in a high-dimensional space, where semantically similar words are located closer to each other. Alnabhan and Branco [10] highlight the transition from handcrafted features to deep learning models that automatically extract hierarchical text representations, facilitating better generalization. Deep contextual models also handle subtleties such as sarcasm, irony, and hidden intent, which traditional classifiers often overlook. Context-based detection, which considers information beyond the text itself, offers another crucial dimension in identifying misinformation.

2.2 Context-Based Detection

Content alone is not sufficient in cases where the text of fake news resembles that of legitimate articles. Context-aware detection systems incorporate metadata such as the publisher's reputation, the sharing user's credibility, network propagation paths, and the temporal sequence of news diffusion. For example, news originating from a source with a history of publishing false information is more likely to be suspect.

Graph Neural Networks (GNNs) have emerged as the most effective solution for modeling relationships among publishers, readers, and news articles. Han et al. [14] demonstrated that GNNs improve fake news detection by learning the underlying structures in user-news engagement graphs, thus capturing social context beyond the article's content. GNNs operate on graph data, where nodes represent entities (e.g., users, articles) and edges represent relationships (e.g., sharing, publishing), allowing the model to learn from the network structure. This shift toward context-based detection recognizes that misinformation spreads differently across social networks compared to authentic news, offering additional discriminatory power. Combining content and context leads to more robust detection systems.

2.3 Hybrid and Multi-Modal Detection

Combining multiple modalities—text, images, metadata, and network graphs—significantly improves detection accuracy. Multi-modal models are capable of validating the relationship between visual and textual information. For example, a misleading image caption can be flagged if the visual features (analyzed via Convolutional Neural Networks (CNNs) [15] or CLIP [16]) contradict the text content (analyzed via BERT). CNNs excel at extracting hierarchical features from images, while CLIP learns joint representations of images and text.

Researchers [17] emphasize on the role of multi-modal approaches in mitigating the limitations of content-only or context-only systems. Such systems are particularly useful for identifying manipulated media and deepfakes, which are increasingly used to spread disinformation, often in highly visual formats on platforms like Twitter, Instagram, and TikTok. Understanding how fake news propagates through networks provides further valuable information.

2.4 Network and Propagation-Based Detection

Propagation-based detection relies on the observation that fake news spreads differently than real news. Specifically, fake news often spreads more virally but has a shorter life span, whereas real news diffuses at a more moderate pace and tends to sustain engagement for longer periods. This behavior has been noted across diverse social platforms such as Twitter, Facebook, and Reddit. Analyzing the cascade patterns of information sharing can reveal anomalies indicative of manipulation.

Han et al. [14] and Mahmud et al. [18] proposed models that capture these behaviors using GNN-based architectures, enabling the detection system to generalize better across varying topics and platforms. A deeper understanding of these spread patterns has enabled both academia and industry to propose hybrid solutions combining social graphs, user embeddings, and temporal patterns. The involvement of human input and the methods used to evaluate these models are also critical aspects.

2.5 Human-in-the-Loop and Evaluation Metrics

Beyond purely automated models, human-in-the-loop systems have also shown promise in improving

accuracy through continuous learning and feedback. Crowdsourcing platforms have allowed datasets to incorporate human judgment at scale, especially for subjective assessments. These systems leverage human expertise to refine model predictions, especially in ambiguous cases.

Evaluation metrics such as precision, recall, F1-score, and Area Under the Receiver Operating Characteristic curve (AUC-ROC) are essential for comparing models fairly, especially when class imbalance is significant [19]. Precision measures the accuracy of positive predictions, while recall measures the ability to identify all positive instances. The F1-score is the harmonic mean of precision and recall, and AUC-ROC provides an overall measure of the classifier's performance across different thresholds. Metrics tailored to real-time detection, like latency and early detection rate, are gaining attention as practical benchmarks in social media settings. The datasets used to train and evaluate these models are the foundation of this research area.

2.6 Temporal Dynamics in Fake News Research

While much of the early work focused on static datasets, there is a growing recognition of the importance of the temporal aspects of fake news. For instance, the analysis of news spread patterns over time has revealed that misinformation often exhibits distinct diffusion characteristics compared to genuine news, including higher initial velocity but potentially shorter lifespan [1]. Furthermore, the fluctuating volume of fake news during significant global events, as visually depicted in Figure 1, underscores the need for detection models that are robust to these temporal shifts and can effectively identify misinformation during crises when its impact can be most damaging. Future research should focus on developing dynamic models that can adapt to evolving patterns of misinformation spread and account for temporal dependencies in both textual and visual cues. This includes exploring techniques like time-series analysis of social media data, incorporating temporal features into detection models, and evaluating model performance across different time periods and during specific events.
technologies and communication mechanisms that lie beneath them, is presented here, along with a discussion of recent developments in both the theoretical underpinnings and practical applications of wireless underground communications. The most significant difficulties encountered in the design and implementation of Ag-IoT are also covered.

III. Datasets for Fake News Detection

Benchmark datasets are vital for model training and evaluation. Key datasets include:

- LIAR [9]: Political fact-checking dataset with short statements and verdict labels, primarily focusing on textual content.
- ISOT [20]: Real and fake news articles on various topics, mainly text-based.
- FakeNewsNet [14]: Combines news articles, social context (user interactions, network structure), and engagement metadata, offering a richer context.
- Fakeddit [18]: Multi-modal Reddit dataset supporting image and text analysis, specifically designed for studying cross-modal interactions.

These datasets vary in size, domain, and the modalities they include, influencing the types of models that can be effectively trained and evaluated on them. The features extracted from these datasets form the input for the detection models.

The increasing trend of fake news proliferation over the years, as shown in Figure 2, highlights the importance of creating and utilizing diverse and representative datasets to train robust detection models. These datasets vary in size, domain, and the modalities they include, influencing the types of models that can be effectively trained and evaluated on them.

IV. Feature Extraction and Models

4.1. Textual Features

Tokenization (breaking text into words), text normalization (e.g., stemming, lemmatization), and embedding methods such as Word2Vec [11], GloVe [12], and BERT [13] allow for efficient and

meaningful vectorization of news articles [19, 20]. Word embeddings capture semantic relationships between words. Paragraph-level and sentence-level embeddings (e.g., Sentence-BERT [21]) further enhance model sensitivity to context shifts, especially for nuanced language typical in misinformation campaigns. Sentence-BERT, for instance, is fine-tuned to produce sentence embeddings that are semantically meaningful and can be compared using cosine similarity.

4.2. Visual and Network Features

Image-based fake news often relies on manipulated visuals. Visual analysis using CNNs [15], along with network feature extraction through GNNs [22], has proven to strengthen the robustness of detection systems [14, 17]. CNNs can automatically learn hierarchical features from images, identifying patterns indicative of manipulation. GNNs, as discussed earlier, model the relationships within social networks. Combining these visual representations with text vectors leads to substantial performance boosts in multi-modal classifiers. The analysis of trends in fake news can provide valuable insights.

V. Visual Analysis of Fake News Trends

Figures 1 and 2, depicting “Fake News Volume Trends across Major Global Events” and the “Distribution of Real vs. Fake News over Years” respectively, offer valuable visual insights into the temporal dynamics and overall prevalence of misinformation. Figure 1 illustrates how the volume of fake news can fluctuate significantly during specific global events, highlighting the challenges for detection systems in handling time-sensitive situations. Similarly, Figure 2 demonstrates the evolving scale of the problem by showing the distribution of real versus fake news over time, underscoring the persistent need for effective detection mechanisms. These visual representations emphasize the dynamic and growing nature of the fake news challenge, reinforcing the importance of developing robust and adaptive detection systems. The complexities involved in building such systems will be explored in the following sections.

VI. Challenges and Future Directions

Fake news detection remains a challenging problem, not only due to the technical aspects of modeling but also because of its sociocultural implications and the rapidly evolving na ture of misinformation tactics. The following subsections provide a detailed examination of persistent research gaps and future opportunities.

6.1. Generalization and Domain Adaptation

Many fake news detection models are trained and evaluated on specific datasets that are limited in domain and context. When these models encounter news from new sources or platforms, their performance often degrades significantly. This is primarily because fake news adapts quickly to social, linguistic, and geopolitical contexts. For example, the surge in fake news during specific global events, as illustrated in Figure 1, highlights the challenge of developing models that can generalize across different situations and time periods. A model trained on political fake news might struggle to identify misinformation in the health domain due to differences in vocabulary and style. The generalization gap highlights the importance of developing adaptive models capable of transferring knowledge across domains, languages, and social dynamics. One promising direction is to incorporate continual learning frameworks that allow models to learn new patterns incrementally without catastrophic forgetting. A key research question here is: *How can continual learning strategies be effectively integrated into multi-modal fake news detection models to maintain performance across diverse and evolving information landscapes?*

6.2. Linguistic and Cultural Diversity

Fake news is a global phenomenon, but most research focuses on English-language data. Low-resource languages, regional dialects, and cultural nuances introduce significant complexity to fake news detection. The lack of multilingual datasets and pre-trained models for these languages hampers

progress. Multilingual language models like XLM RoBERTa [23] show potential, but even these struggle with deep cultural context. For instance, detecting sarcasm or irony can be highly culture-specific. Future research must prioritize the creation of high-quality datasets for underrepresented languages and the development of culture-aware fake news classifiers. A crucial research question is: *What novel techniques can be developed to effectively capture and leverage linguistic and cultural nuances for fake news detection in low-resource languages?*

6.3. Explainability and Transparency

Modern deep learning architectures such as transformers [24] and graph neural net works [22] are often criticized for their lack of interpretability. For real-world deployment, especially in journalism, policy enforcement, and legal proceedings, models must offer clear explanations for their decisions. Explainable AI (XAI) techniques—such as attention visualization, counterfactual reasoning, and feature importance extraction—should be integrated into fake news detection systems to enhance trust and accountability. For example, highlighting the specific words or image regions that contributed most to the model's prediction can provide valuable insights. A key research question is: *How can XAI methods be effectively adapted and applied to multi-modal fake news detection to provide meaningful and trustworthy explanations for model predictions?*

6.4. Data Imbalance and Label Noise

A major hurdle in training reliable fake news detectors is the imbalance in datasets—authentic news heavily outnumbers fake samples. This imbalance can cause the models to be biased towards real news, reducing sensitivity to false information. Furthermore, mislabeled data due to subjective annotation or evolving definitions of fake news introduce noise during training. Solutions like adversarial training [25], ensemble learning [26], data augmen tation, and synthetic data generation using techniques such as Generative Adversarial Networks (GANs) [25] offer avenues for improving performance in the presence of imbal ance and noise. A critical research question is: *What are the most effective strategies for addressing data imbalance and label noise in the context of multi-modal fake news detection to improve model robustness and generalization?*

6.5. Human-Centered Design and Ethical Considerations

Beyond algorithmic advancements,the detection of fake news must consider human interaction. Systems that explain their decisions in plain language and allow user feedback loops will be more effective in practice. Additionally, ethical considerations such as censorship, bias in training data (e.g., models trained primarily on one political viewpoint might unfairly flag news from another), and the potential misuse of fake news detection tools (e.g., to suppress legitimate but dissenting opinions) must be addressed during system design. A crucial research question is: *How can human-centered design principles and ethical frameworks be integrated into the development and deployment of multi modal fake news detection systems to ensure fairness, transparency, and accountability?*

VII. Conclusion

In conclusion, the pervasive threat of fake news to the integrity of public discourse demands continued and concerted attention. The evolving landscape of digital communication and content creation necessitates a dynamic shift in detection strategies. As we have explored, the future of effective misinformation detection lies in the convergence of multi modal analysis, the integration of explainable AI for enhanced trust and accountability, and the development of continually learning frameworks capable of adapting to novel deceptive tactics. Addressing the inherent challenges of generalization, linguistic diversity, data scarcity, and ethical considerations will be paramount. Ultimately, progress in this critical field hinges on robust interdisciplinary collaboration, uniting expertise from computational science, journalism, social sciences, and policy to forge detection systems that are not only accurate and reliable but also ethically sound and resilient against the ever-evolving tide of online deception.

References

[1] S. Vosoughi, D. Roy, and S. Aral, ”The spread of true and false news online,” Science, vol. 359, no. 6380, pp. 1146–1151, 2018.
[2] K. C. Shu, D. Wang, and H. Liu, ”Understanding and detecting misinformation in online social media,” Synthesis Lectures on Human-Centered Informatics, vol. 13, no. 1, pp. 1–173, 2020.
[3] M. C. van Zuijlen, ”The platformization of news: Mapping the infrastructure of social media news,” Social Media + Society, vol. 6, no. 1, 2020.
[4] H. Allcott and M. Gentzkow, ”Social media and fake news in the 2016 election,” Journal of Economic Perspectives, vol. 31, no. 2, pp. 211–236, 2017.
[5] E. C. Tandoc Jr, Z. W. Lim, and R. Ling, ”Defining “fake news”: A typology of scholarly definitions,” Digital Journalism, vol. 6, no. 2, pp. 137–153, 2018.
[6] K. C. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, ”Fake news detection on social media: A data mining perspective,” ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.
[7] Z. Jin, J. Cao, Y. Zhang, and J. Luo, ”News veracity identification by cross-modal fusion,” in Proceedings of the 25th ACM international conference on Multimedia, pp. 1688–1691, 2017.
[8] C. J. Vargo, L. Guo, and M. K. Amazeen, ”The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016,” New Media and Society, vol. 20, no. 7, pp. 2029–2049, 2018.
[9] X. Zhou and R. Zafarani, ”A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities,” ACM Computing Surveys, vol. 53, no. 5, pp. 1–40, 2020.
[10] M. Q. Alnabhan and P. Branco, ”Fake News Detection Using Deep Learning: A Systematic Literature Review,” IEEE Access, vol. 11, pp. 12540–12563, 2023.
[11] T. Mikolov, K. Chen, G. Corrado, and J. Dean, ”Efficient estimation of word representations in vector space,” arXiv preprint arXiv:1301.3781, 2013.
[12] J. Pennington, R. Socher, and C. D. Manning, ”Glove: Global vectors for word representation,” in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532–1543, 2014.
[13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, ”Bert: Pre-training of deep bidirectional transformers for language understanding,” arXiv preprint arXiv:1810.04805, 2018.
[14] Y. Han, S. Karunasekera, and C. Leckie, ”Graph Neural Networks with Continual Learning for Fake News Detection from Social Media,” arXiv preprint arXiv:2007.03316, 2020.
[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ”Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, pp. 1097–1105, 2012.
[16] A. Radford et al., ”Learning transferable visual models from natural language supervision,” in International conference on machine learning, pp. 8748–8763, PMLR, 2021.
[17] Y. Mohan and C. Kanwar, ”A Review of NLP And ML Algorithms For The Detection of Fake News on Social Media,” IJSART, vol. 9, no. 7, 2023.
[18] T. Mahmud, T. Akter, M. T. Aziz, et al., ”Integration of NLP and Deep Learning for Automated Fake News Detection,” ICICI-2024 Conference Proceedings, 2024.
[19] M. Madani, H. Motameni, and R. Roshani, ”Fake News Detection Using Feature Ex traction, Natural Language Processing, Curriculum Learning, and Deep Learning,” 2023.
[20] H. F. Villela, F. Corrˆea, J. S. A. N. Ribeiro, and A. Rabelo, ”Fake news detection: A systematic literature review of machine learning algorithms and datasets,” 2023.
[21] N. Reimers and I. Gurevych, ”Sentence-bert: Sentence embeddings using siamese bert-networks,” arXiv preprint arXiv:1908.10084, 2019.
[22] T. N. Kipf and M. Welling, ”Semi-supervised classification with graph convolutional networks,” arXiv preprint arXiv:1609.02907, 2017.
[23] A. Conneau et al., ”Unsupervised cross-lingual representation learning at scale,” in

International conference on learning representations, 2020.
[24] A. Vaswani et al., ”Attention is all you need,” in Advances in neural information processing systems, pp. 5998–6008, 2017.
[25] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, ”Generative adversarial nets,” in Advances in neural information processing systems, pp. 2672–2680, 2014.
[26] T. G. Dietterich, ”Ensemble methods in machine learning,” in International workshop on multiple classifier systems, pp. 1–15, Springer, 2000.