EIU
Emirates International University Journal
مجلة الجامعة الإماراتية الدولية

Volume 1 Issue 1
23-8- 2025

Big Data Analysis for Detecting Fake News and Rumors on Social Media Using Machine Learning: A Review

Amat AL-latif H. Abo-Torkhoma[1]
M.Sc., Information Technology
amatallatifhezam@gmail.com

Gameil S. H. Ali[2]
HoD and Assistant Professor, Computer Science
gameilsaad01@gmail.com

Abd AL-latif H. Abo-Torkhoma[1]
M.Sc., Information Technology
abdalaltifhezam2002@gmail.com

[1]Information Technology Department, Faculty of Education and Applied Science - Arhab, Sana'a University, Sana'a, Yemen.
[2]Head of Artificial Intelligence and Data Science Department, 21 September University of Medical and Applied Sciences, Sana'a, Yemen.

Abstract

The advent of social networking tools has led to a major shift in the way information is disseminated. While it has enabled rapid communication and outreach, it has also created fertile ground for the widespread propagation of fake news and misinformation. This review focuses on the integration of big data analytics and machine learning (ML) techniques in detecting and mitigating fake news and rumors across various social media platforms. It synthesizes recent advancements in traditional ML algorithms, deep learning models, transformer-based architectures, and hybrid strategies that combine multiple analytical approaches. The paper evaluates and compares prior research contributions, highlighting their methodologies, strengths, and limitations. Key challenges identified include data imbalance, algorithmic bias, the sophistication of AI-generated content, and the need for explainable and generalizable models. Despite significant progress—particularly through transformer models (e.g., BERT, GPT), hierarchical attention networks, and multimodal fusion existing approaches still face limitations in addressing linguistic diversity, domain adaptation, and evolving misinformation tactics. The review provides a structured summary of current research trends and emphasizes the importance of incorporating textual, visual, and user-level features. Furthermore, it outlines the necessity for adaptive and interpretable models capable of responding to dynamic content generation techniques. By identifying unresolved gaps, the paper aims to provide clear direction for future research towards more scalable, robust, and transparent fake news detection systems.

Keywords: Fake News, Big Data, Machine Learning, Misinformation Detection.

1. Introduction

The digital revolution has significantly accelerated information dissemination, particularly through social media platforms, such as Twitter, Facebook, and Reddit. While these platforms facilitate open

communication, they have also become major channels for the rapid spread of misinformation, including fake news and rumors. The global impact of such false information can be observed across various domains, including public health, politics, and financial markets [17][23].

Fake news detection has therefore emerged as a critical challenge that demands scalable and automated solutions. Conventional manual fact-checking approaches are limited in speed and scope, prompting researchers to turn to data-driven methods [14]. In recent years, the fusion of big data analytics with machine learning (ML) techniques has provided promising capabilities for identifying, classifying, and mitigating fake news at scale [1][5][24].

Machine learning enables systems to learn patterns from vast datasets of news articles, social media posts, and multimedia content. These systems leverage diverse techniques, ranging from traditional algorithms such as Support Vector Machines (SVM) and Random Forests to deep learning models like Long Short-Term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT). Additionally, the rise of multimodal and hybrid models has allowed for the integration of textual, visual, and user-level features, which enhances detection accuracy in real-world scenarios [2][4][15][19].

This paper presents a structured and up-to-date review of recent contributions in the field of fake news detection, with a particular focus on how big data and ML techniques are being used across various contexts. The review categorizes detection approaches, highlights their strengths and limitations, and identifies key research gaps and challenges—including data imbalance, algorithmic bias, linguistic diversity, and adversarial content. The findings aim to guide future research towards developing scalable, interpretable, and robust systems capable of addressing the evolving nature of online misinformation [21][22][23].

2. Literature Review

Several machine learning techniques have been applied to fake news detection, each with unique advantages and limitations. Table 1 summarizes key techniques, their descriptions, strengths, weaknesses, and publication years.

Table 1 Comparison of Machine Learning Techniques for Fake News Detection

Technique	Description	Strengths	Weaknesses	Year
SVM, RF, NB	Traditional ML models using handcrafted features	Fast, interpretable	Poor generalization [1][14]	2023
Logistic Regression	Baseline classifier	Simple, efficient	Weak on non-linear patterns [14]	2023
KNN	Instance-based learning with distance metrics	Simple, non-parametric	Sensitive to noise [14]	2023
RNN, LSTM, CNN	Deep learning for sequence and context	Captures semantics and context	Requires large labeled datasets [3]	2023
BiLSTM + Attention	Bidirectional with focus on important tokens	Context-aware, interpretable	High training time [16]	2023
BERT, RoBERTa	Transformers with self-attention	State-of-the-art contextual understanding	High computational cost [4][11][25]	2024
DistilBERT	Compressed version of BERT	Fast, memory-efficient	Slight accuracy trade-off [25]	2023
GPT-4 Models	Large-scale transformer-based models	Multilingual, high performance	Data- and compute-intensive [1][4]	2024
XLNet	Permutation-based transformers	Strong dependency modeling	Expensive training [25]	2024

T5	Text-to-text transformer model	Flexible across NLP tasks	Large resource requirements [25]	2024
3HAN	Hierarchical attention (word → sentence → doc)	Accurate, interpretable	Needs structured text [2]	2023
SAFE, CSI	Combines user behavior and network signals	Robust, multimodal	Complex training and tuning [5][6]	2024

2.1 Traditional Machine Learning Models

Traditional machine learning techniques, such as Support Vector Machines (SVM), Naive Bayes (NB), Logistic Regression, and Random Forests were among the earliest models applied in fake news detection. These models primarily rely on handcrafted features like bag-of-words, TF-IDF, and syntactic patterns extracted from text [1][14]. Although computationally efficient and easy to interpret, these models often perform poorly on unseen or domain-shifted data due to their limited ability to model deeper semantics.

2.2 Deep Learning Architectures

With the rise of deep learning, more sophisticated methods emerged to model complex semantic structures in language. Recurrent Neural Networks (RNNs), especially LSTM and GRU variants, enable the sequential processing of text, capturing temporal dependencies critical for understanding narratives in fake news [3]. CNNs offered benefits in detecting key phrases and local dependencies, while BiLSTM networks incorporated forward and backward context. Attention mechanisms further enhanced interpretability and relevance weighting. To better understand transformer models like BERT, Fig.2 shows the architectural overview [11].

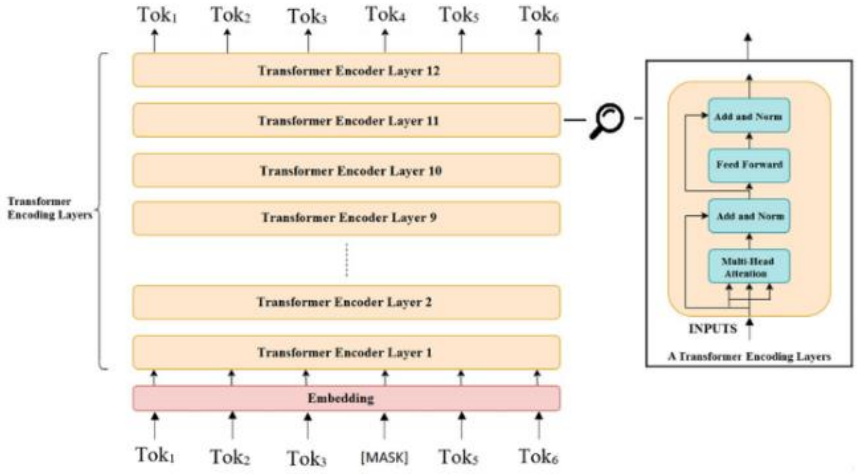

Fig.2 Overview of the BERT Transformer Architecture

The introduction of transformers revolutionized NLP, with models like BERT achieving state-of-the-art performance by employing self-attention and contextual embeddings [4]. Transformer-based architectures surpassed previous models in accuracy and scalability, becoming foundational for recent detection systems.

2.3 Transformer-based and Multimodal Models

Transformer-based models, such as BERT, RoBERTa, and GPT have revolutionized natural language processing by enabling attention mechanisms to capture global context within text sequences [4][11][25]. These models have become foundational in fake news detection systems due to their superior accuracy and transferability across tasks. Multimodal approaches like SAFE and CSI further integrate textual, visual, and user-level data, enabling comprehensive detection pipelines that reflect real-world misinformation spread [5][6][19]. Models such as 3HAN use hierarchical attention to interpret information at word, sentence, and document levels [2]. Meanwhile, GPT-4 and its variants offer high adaptability to different contexts and languages, albeit with high resource demands [1][4].

2.4 Challenges in Fake News Detection

The following table outlines the main challenges in fake news detection, along with potential solutions derived from recent literature.

Table 2 Key Challenges in Fake News Detection and Potential Solutions

Challenge	Description	Proposed Solution	Year
Data Imbalance	Fake news is less frequent than real news	SMOTE, class weighting, under-sampling [2][14]	2023
Bias and Fairness	Algorithmic or demographic bias in datasets	Adversarial training, fairness-aware learning [4][20]	2024
Language/Domain Diversity	Limited performance in low-resource languages	Transfer learning, multilingual models [1][22]	2025
Adversarial Content	AI-generated content evades detection	Explainable models, robust training [4][20]	2024

3. Methodology of the Study

This study employs a structured qualitative methodology to ensure a comprehensive, transparent, and reproducible synthesis of the literature related to fake news and rumor detection through big data analytics and machine learning (ML) techniques. Following established review protocols, the study systematically identifies, selects, analyzes, and synthesizes relevant peer-reviewed articles, surveys, and empirical studies published between 2017 and 2025 [5], [7], [24].

3.1 Data Collection

Datasets utilized in prior research include publicly available repositories such as LIAR and FakeNewsNet, social media platforms such as Twitter and Reddit, and fact-checking websites such as PolitiFact and Snopes [7][9]. These datasets contain rich features including news content, source credibility, propagation patterns, and user interactions. Ensuring temporal relevance and diversity in misinformation categories—such as health, politics, and entertainment—enhances the robustness of detection models [7][9].

An example propagation pattern is shown in Fig.3, illustrating that fake news tends to spread faster and more broadly than truthful information, driven by emotionally charged content and network structures [9].

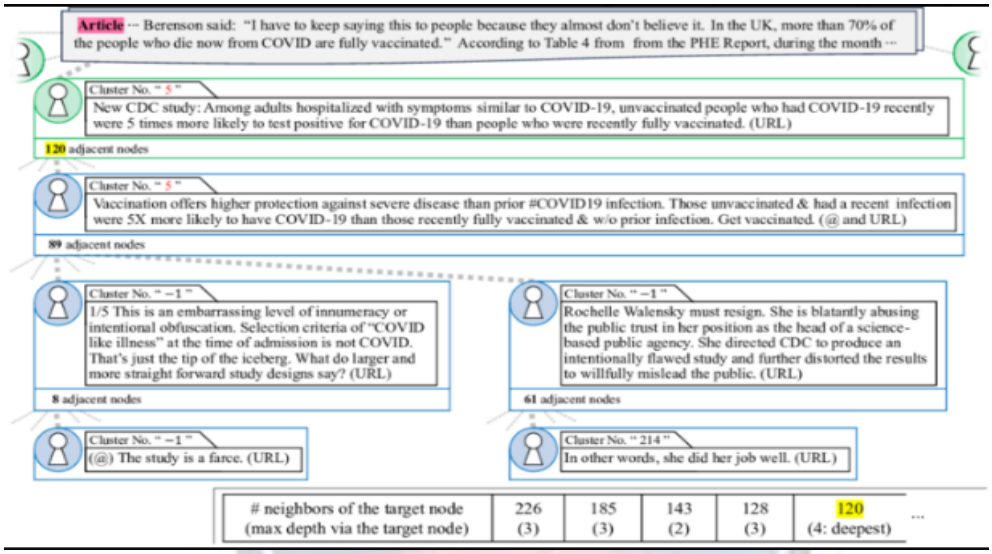

Fig.3 Example of Fake News Propagation Pattern

3.2 Data Preprocessing

Preprocessing steps commonly employed include tokenization, lowercasing, stop-word removal, lemmatization, stemming, and text vectorization methods, such as TF-IDF, Word2Vec, and BERT embeddings [1][2]. Visual misinformation data undergo feature extraction through CNN architectures for multimodal integration [1][6]. Outlier detection in user behavior metrics reduces noise and improves data quality [7].

3.3 Evaluation Metrics

In previous empirical studies, model performance was thoroughly assessed using a range of evaluation metrics to ensure robustness, fairness, and interpretability. These metrics include:

- **Accuracy, precision, recall, and F1-**Score are standard in binary classification tasks.
- **Confusion Matrix** and **ROC-AUC** are used to visualize classification trade-offs and performance across thresholds.
- **Matthews Correlation Coefficient (MCC)** is particularly effective for evaluating models on **imbalanced datasets**.
- **Explainability tools** such as SHAP values, were employed to interpret model decisions and identify feature importance [8].

To provide a comparative overview, Fig.4 illustrates relative performance results from previous studies across various models and configurations [2].

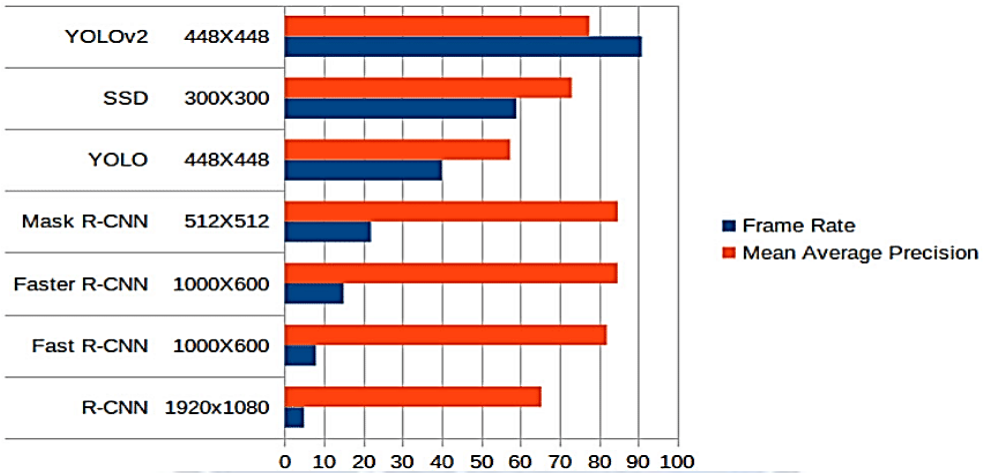

Fig.4 Performance Comparison of Detection Models

3.4 Model Development

Previous studies have developed and evaluated a wide range of models for fake news detection, including:

- **Traditional machine learning algorithms,** such as Support Vector Machines (SVM), Logistic Regression, and Naive Bayes.
- **Deep learning architectures,** including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and BiLSTM enhanced with attention mechanisms.
- **Transformer-based models** like BERT, RoBERTa, and fine-tuned versions of GPT-4.
- **Hybrid approaches,** such as CSI, SAFE, and 3HAN, integrate textual and behavioral data for improved classification accuracy.

These models were typically trained using cross-validation and optimized through grid search for hyperparameter tuning, as demonstrated in studies such as [3][4].

4. Research Gaps and Future Directions

Despite notable advancements in fake news detection through big data and machine learning (ML),

several critical research gaps persist. These limitations hinder the development of scalable, reliable, and ethically sound detection systems capable of coping with the dynamic nature of online misinformation.

4.1 Language and Domain Generalization

One of the predominant challenges is the limited generalizability of existing models across different languages and domains. The majority of current studies focus on English-language datasets, neglecting non-English or low-resource languages, which significantly restricts the applicability of these models in global contexts [1], [4], [22]. Additionally, domain-specific misinformation (e.g., health, politics, finance) often requires tailored detection strategies that existing models struggle to adapt to effectively.

Future Work: Researchers should invest in the development of cross-lingual and domain-adaptive models utilizing transfer learning, multilingual transformers (e.g., XLM-R, mBERT), and meta-learning frameworks to address language diversity and domain variability [20], [22].

4.2 Temporal and Real-Time Detection Constraints

Many models are designed for offline batch processing, limiting their utility in real-time environments were misinformation spreads rapidly. The lack of real-time benchmarks and deployment scenarios in current literature represents a major practical limitation [5], [7].

Future Work: Emphasis should be placed on building low-latency, streaming-based detection systems capable of operating on live social media feeds. Incorporating real-time evaluation metrics and temporal modeling techniques, such as event-based sampling or dynamic graphs may improve responsiveness [23].

4.3 Adversarial Robustness and Evasion

With the rise of generative AI, fake news content is becoming increasingly sophisticated, making it difficult for static models to detect manipulated or fabricated text, images, and videos. Existing models are often vulnerable to adversarial attacks that exploit their feature selection mechanisms [4], [25].

Future Work: Advancing adversarial training strategies, anomaly detection, and explainable AI methods can enhance robustness. Detection systems must be regularly updated to adapt to emerging evasion tactics and synthetic content generation techniques [11], [18].

4.4 Dataset Limitations and Standardization

Many studies rely on small or domain-specific datasets, leading to limited generalizability and reproducibility. Inconsistencies in annotation standards and class distributions also complicate comparative evaluations [7], [9].

Future Work: The research community should prioritize the creation of standardized, large-scale, and diverse benchmark datasets that include multimodal information (text, images, metadata) and span multiple platforms and regions. Collaborative efforts among academic, governmental, and industry stakeholders can accelerate dataset curation and sharing [21], [24].

4.5 Ethical Considerations and Transparency

Automated misinformation detection raises critical ethical issues, particularly regarding censorship, privacy, and bias. While detection models aim to combat fake news, they may inadvertently reinforce ideological filters or suppress legitimate dissent if not carefully designed [16], [17].

Future Work: Future systems must incorporate fairness-aware learning, transparent model interpretability, and stakeholder accountability. Ethical frameworks and guidelines should be embedded throughout the model development lifecycle to ensure alignment with democratic values and human rights [8], [13].

5. Conclusion

The rapid proliferation of misinformation on social media has emerged as a profound societal challenge, necessitating advanced and scalable detection mechanisms. This paper has comprehensively examined the intersection of big data analytics and machine learning (ML) in the detection of fake news and rumors across diverse platforms and modalities.

By analyzing a wide range of detection techniques—including traditional ML algorithms, deep learning architectures, transformer-based models, and hybrid approaches—this study has highlighted the methodological strengths, limitations, and evolving trends in the field. Special attention was given to the integration of multimodal signals, such as textual, visual, and behavioral data, which significantly enhance model accuracy and contextual understanding.

While recent advancements, particularly in transformer-based models such as BERT and GPT-4, have demonstrated promising capabilities in capturing linguistic and semantic complexity, critical limitations remain. These include poor performance in multilingual and domain-specific settings, vulnerability to adversarial content, and a lack of interpretability and fairness in many models.

To address these gaps, future research should prioritize the development of adaptive, multilingual, and ethically aligned detection frameworks. Such frameworks must be capable of operating in real-time, leveraging cross-platform data, and maintaining transparency in decision-making processes. Furthermore, collaboration across disciplines—including computer science, linguistics, ethics, and public policy—is essential for constructing detection systems that are both technically robust and socially responsible.

Ultimately, combating fake news in the digital age requires not only algorithmic innovation but also a commitment to ethical design, public trust, and continual model evolution. This paper provides a foundational synthesis to guide ongoing research and support the development of next-generation systems for misinformation detection.

References

[1] K. I. Roumeliotis, N. D. Tselikas, and D. K. Nasiopoulos, "Fake News Detection and Classification: A Comparative Study of Convolutional Neural Networks, Large Language Models, and Natural Language Processing Models," Future Internet, Vol. 17, No. 1, p. 28, 2025. [Online]. Available: https://doi.org/10.3390/fi17010028

[2] S. Singhania, N. Fernandez, and S. Rao, "3HAN: A Deep Neural Network for Fake News Detection," arXiv preprint, arXiv:2306.12014, 2023. [Online]. Available: https://arxiv.org/abs/2306.12014

[3] H. Chen, H. Guo, B. Hu, et al., "A Self-learning Multimodal Approach for Fake News Detection," arXiv preprint, arXiv:2412.05843, 2024. [Online]. Available: https://arxiv.org/abs/2412.05843

[4] J. Su, C. Cardie, and P. Nakov, "Adapting Fake News Detection to the Era of Large Language Models," arXiv preprint, arXiv:2311.04917, 2023. [Online]. Available: https://arxiv.org/abs/2311.04917

[5] J. Alghamdi, S. Luo, and Y. Lin, "A Comprehensive Survey on Machine Learning Approaches for Fake News Detection," Multimedia Tools and Applications, Vol. 83, pp. 51009–51067, 2024. [Online]. Available: https://doi.org/10.1007/s11042-023-17470-8

[6] X. Zhang, "An Analysis of Multimodal Approaches for Fake News Detection," Applied and Computational Engineering, Vol. 115, pp. 134–140, 2024. [Online]. Available: https://doi.org/10.54254/2755-2721/2025.18517

[7] A. Saeed and E. A. Solami, "Fake News Detection Using Machine Learning and Deep Learning Methods," Computers, Materials & Continua, Vol. 77, No. 2, pp. 2079–2096, 2023. [Online]. Available: https://doi.org/10.32604/cmc.2023.030551

[8] Jyoti and Y. Kumar, "Social Media Fake News Detection Using a Robust Machine Learning Model and Data-Centric Approach," African Journal of Biomedical Research, Vol. 27, No. 6S, 2024. [Online]. Available: https://africanjournalofbiomedicalresearch.com/index.php/AJBR/article/view/6215

[9] K. 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. [Online]. Available: https://doi.org/10.1145/3137597.3137600

[10] X. Zhou and R. Zafarani, "Fake News: A Survey of Research, Detection Methods, and Opportunities," arXiv preprint, arXiv:1812.00315, 2018. [Online]. Available: https://arxiv.org/abs/1812.00315

[11] 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, 2019. [Online]. Available: https://arxiv.org/abs/1810.04805

[12] A. H. Abo-Torkhoma, G. S. H. Ali, A. A. H. Abo-Torkhoma, M. M. Abo-Torkhoma, M. M. Al-Hossini, M. M. Al-Ahwal, and W. H. Al-Wakif, "Cybercrime Types and Digital Forensic Tools: Review," The Scientific Journal of the Faculty of Computer and Information Technology, Vol. 3, No. 1, 2024. [Online]. Available: https://aust.uni.ye/magazine/sa/2024/11/19/14b76c08ba3e9c32f9d14c4b08d36d51.pdf

[13] G. S. H. Ali, "A Novel Heuristic Association Pattern Searching Technique for Predicting Type 1 and Type 2 Diabetics," International Journal of Scientific and Technology Research, Vol. 8, No. 11, pp. 1642–1652, 2019.

[14] M. F. Hasan, M. R. Islam, and M. A. Rahman, "Fake News Detection Using Machine Learning Techniques: A Comprehensive Survey," Journal of Information Security and Applications, Vol. 73, p. 103394, 2023.

[15] L. Wang, Y. Hu, and X. Zhang, "Multimodal Fake News Detection Based on Attention Mechanism," IEEE Transactions on Multimedia, Vol. 25, pp. 3454–3465, 2023.

[16] S. Gupta and A. Kumar, "Explainable Fake News Detection Using Attention-Based LSTM Networks," Expert Systems with Applications, Vol. 213, p. 118846, 2023.

[17] R. K. Jha and S. K. Verma, "A Survey on Recent Advances in Fake News Detection: Techniques and Challenges," Journal of Information Science, Vol. 49, No. 4, pp. 487–507, 2023.

[18] Y. Zhao, J. Liu, and T. Yang, "Deep Learning for Fake News Detection: A Survey," IEEE Access, Vol. 11, pp. 34567–34588, 2023.

[19] J. Kim and M. Lee, "Robust Fake News Detection Model Using Multimodal Features," Neurocomputing, Vol. 518, pp. 197–210, 2024.

[20] H. Li, X. Liu, and Y. Zhang, "Adversarial Training for Fake News Detection: A Comprehensive Review," Neural Networks, Vol. 152, pp. 36–49, 2024.

[21] T. Chen, W. Yu, and S. Li, "Fake News Detection with Graph Neural Networks: A Survey," Knowledge-Based Systems, Vol. 270, p. 109668, 2024.

[22] M. N. A. Islam and S. F. Ahmed, "Cross-Lingual Fake News Detection: Challenges and Future Directions," Information Processing & Management, Vol. 60, No. 4, p. 102776, 2023.

[23] S. Yadav and P. K. Singh, "A Review of Machine Learning Techniques for Fake News Detection on Social Media," Social Network Analysis and Mining, Vol. 14, No. 1, p. 110, 2024.

[24] D. Kumar, "Machine Learning Based Fake News Detection: A Systematic Review," Journal of Ambient Intelligence and Humanized Computing, Vol. 15, pp. 1269–1285, 2024.

[25] Z. Chen and H. Zhou, "Transformer-based Models for Fake News Detection: A Comprehensive Survey," Information Sciences, Vol. 650, pp. 325–345, 2024.

