International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 15, Issue 2, February 2026)

Machine Learning Techniques for Fake News Detection: A Review

[1]Neha Sharma, [2]Prof. (Dr.) Vishal Kohli
[1]Research Scholar, Department of Computer Science & Engineering, Neelkanth Institute of Technology, Meerut, India
[2]Director, Department of Computer Science & Engineering, Neelkanth Institute of Technology, Meerut, India

***Abstract*— A Review examines how artificial intelligence methods are used to identify and classify misleading information circulated through digital platforms. The rapid growth of social media and online news portals has increased the spread of fabricated content, creating serious social, political, and economic challenges. This review summarizes various machine learning approaches such as Naïve Bayes, Support Vector Machine (SVM), Decision Tree, Random Forest, K-Nearest Neighbor, and deep learning models including Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks for fake news classification. It discusses commonly used datasets, feature extraction methods like TF-IDF, word embeddings, and n-grams, as well as performance evaluation metrics such as accuracy, precision, recall, and F1-score. The study also highlights current challenges including data imbalance, multilingual content, and evolving misinformation patterns, and outlines future research directions focusing on hybrid models, explainable AI, and real-time detection systems.**

***Keywords*—AI Fake News Detection, ML, DL, Text Classification, Social Media Analytics**

I. Introduction

Fake news detection has become one of the most important research areas in the fields of Artificial Intelligence, Data Science, and Cybersecurity. In the digital age, information spreads rapidly through online platforms such as Facebook, Twitter, Instagram, and WhatsApp[1]. While these platforms provide fast communication and global connectivity, they also enable the rapid spread of misleading, false, or manipulated information. Fake news refers to fabricated or intentionally misleading content presented as real news, often created to influence public opinion, gain financial benefits, or create social and political instability[2].

The problem of fake news is not new, but its scale has significantly increased due to the widespread use of the internet and smartphones. In earlier times, traditional media such as newspapers and television had strict editorial processes to verify information before publication[3]. However, with the rise of user-generated content and digital journalism, anyone can publish information without proper verification. This has created an environment where rumors, propaganda, and misinformation can quickly reach millions of users within minutes[4].

Fake news can appear in different forms, including fabricated stories, manipulated images, misleading headlines, deepfake videos, and partially true information taken out of context. It is commonly spread during elections, natural disasters, health emergencies, and social conflicts[5]. For example, during global health crises such as the COVID-19 pandemic, false medical advice and conspiracy theories spread widely online, causing confusion and public panic. Similarly, political misinformation has influenced voter behavior in several countries, raising serious concerns about democracy and public trust[6].

Detecting fake news manually is very difficult due to the massive volume of content generated every second. Millions of posts, articles, tweets, and videos are uploaded daily. Human fact-checkers cannot efficiently analyze all this information in real time. Therefore, automated systems based on ML and DL techniques have become essential for identifying and filtering fake content. These intelligent systems analyze patterns in text, user behavior, linguistic style, and source credibility to classify news as real or fake[7].

Machine Learning techniques such as Naïve Bayes, Support Vector Machine (SVM), Decision Tree, Random Forest, and K-Nearest Neighbor (KNN) have been widely used for text classification tasks, including fake news detection. These models rely on feature extraction methods such as Term Frequency–Inverse Document Frequency (TF-IDF), Bag-of-Words (BoW), and n-grams to convert textual information into numerical form[8].

In recent years, the integration of Natural Language Processing (NLP) techniques with Deep Learning has further enhanced detection accuracy. Word embedding methods such as Word2Vec, GloVe, and contextual embeddings from transformer-based architectures have improved the understanding of language patterns. Moreover, hybrid models combining machine learning with sentiment analysis, stance detection, and social network analysis are being explored to improve reliability.[9]

Despite significant progress, fake news detection still faces many challenges. One major issue is the dynamic and evolving nature of misinformation. Fake news creators continuously modify writing styles and content strategies to avoid detection. Another challenge is the availability of balanced and high-quality datasets. Many datasets are domain-specific and may not generalize well across languages, cultures, or regions. Additionally, sarcasm, humor, and ambiguous language make it difficult for automated systems to accurately interpret intent[10].

The rise of deepfake technology and AI-generated content has added a new layer of complexity. Advanced generative models can create highly realistic text, images, and videos that are difficult to distinguish from authentic content. This increases the need for more sophisticated detection techniques that combine textual, visual, and contextual analysis[11].

Fake news detection is not only a technical problem but also a social and ethical issue. The implementation of automated filtering systems must balance misinformation control with freedom of expression. Over-filtering may suppress legitimate opinions, while under-filtering may allow harmful content to spread. Therefore, explainable AI (XAI) techniques are becoming important to ensure transparency and trust in detection systems [12].

II. LITERATURE SURVEY

Mishra et al., [1] presented a Graph Neural Network (GNN) based framework for efficient fake news detection. The study focused on modeling relationships between users, posts, and content sources using graph structures. By capturing social interactions and propagation patterns, the model improved contextual understanding of misinformation spread. The authors implemented node embedding and message passing mechanisms for better feature learning. Experimental results showed higher accuracy compared to traditional ML models. The approach demonstrated improved precision and recall values above 97%. The study highlighted the importance of structural information in fake news identification.

Ahmad et al., [2] introduced an enhanced deep learning model for efficient fake news detection. Their framework integrated convolutional layers with attention mechanisms to capture semantic and contextual patterns. The model was trained on benchmark datasets with balanced real and fake samples. Feature engineering included word embeddings and TF-IDF representations. The enhanced model achieved an accuracy of approximately 98%, outperforming baseline CNN and LSTM models. The study emphasized reduced computational complexity and faster convergence. It concluded that hybrid deep architectures significantly improve detection performance.

Rocha et al., [3] conducted a systematic review analyzing the impact of fake news on social media during the COVID-19 pandemic. The study examined misinformation related to health, vaccines, and preventive measures. It highlighted the psychological and social consequences of false information. The authors reviewed multiple detection strategies and fact-checking frameworks. Findings showed that misinformation increased public confusion and anxiety levels. The paper stressed the need for automated AI-based verification tools. It provided insights into public health risks associated with fake content.

Almeida et al., [4] explored the relationship between smartphones, social networks, and fake news from an institutional economics perspective. The study discussed how digital platforms influence user decision-making processes. It analyzed the economic incentives behind misinformation creation and sharing. The authors emphasized behavioral biases and information asymmetry. The research suggested policy-level interventions and platform regulations. It also recommended integrating AI-based monitoring systems. The study contributed to understanding fake news from a socio-economic viewpoint.

Tasnim et al., [5] examined the impact of rumors and misinformation during the COVID-19 crisis. The research focused on how social media amplified unverified health-related claims. It identified patterns of rapid rumor spread and emotional reactions. The study highlighted the negative influence on public behavior and health safety measures. The authors suggested the use of automated content filtering mechanisms. They emphasized collaboration between health authorities and digital platforms. The paper underscored the urgent need for reliable misinformation detection systems.

Phan et al., [6] presented a survey on graph neural network methods for fake news detection. The authors reviewed various GNN architectures, including GCN and GAT models. The survey discussed how graph-based approaches capture propagation structures and user interactions. It compared performance metrics across multiple benchmark datasets. Results indicated that graph-based models achieved accuracy above 96% in several cases. The study also identified challenges such as scalability and dynamic graph updates. It concluded that GNNs offer strong potential for misinformation analysis.

Oshikawa et al., [7] provided a comprehensive survey on natural language processing techniques for fake news detection. The study categorized methods into content-based, context-based, and hybrid approaches. It discussed linguistic feature extraction methods such as n-grams and semantic embeddings. The authors compared traditional ML models with deep learning frameworks. The review highlighted that deep learning models improved contextual understanding. It also discussed dataset limitations and evaluation challenges. The paper served as a foundational reference for NLP-based detection research.

Manzoor et al., [8] conducted a systematic review of machine learning approaches for fake news detection. The study evaluated algorithms such as Naïve Bayes, SVM, Random Forest, and KNN. Comparative analysis showed that SVM and Random Forest achieved higher classification accuracy around 94–96%. The paper emphasized the importance of preprocessing and feature selection. It also discussed challenges related to multilingual datasets. The authors suggested combining ML models with deep learning for better performance. The study provided structured insights into algorithm efficiency.

Li et al., [9] presented a brief survey on deep learning models for fake news detection. The study analyzed CNN, RNN, and LSTM-based frameworks. It highlighted the advantage of automatic feature extraction in deep learning. Experimental comparisons indicated that LSTM models performed better for sequential text analysis. Reported accuracy values exceeded 97% on benchmark datasets. The authors discussed computational costs and training time. The paper concluded that hybrid deep models are promising for future research.

Wu et al., [10] provided a comprehensive survey on graph neural networks, explaining their theoretical foundations and applications. Although not limited to fake news detection, the study described how GNNs process relational data. It discussed spectral and spatial graph convolution techniques. The paper reviewed performance improvements across multiple domains. The survey highlighted scalability challenges and optimization strategies. It emphasized the adaptability of GNNs for misinformation detection tasks. The study offered strong theoretical support for graph-based fake news models.

Pandey et al., [11] conducted a comparative study of Random Forest, SVM, and Naïve Bayes for sentiment analysis optimization. Though focused on sentiment analysis, the findings are relevant for fake news detection tasks. The research highlighted the importance of ensemble learning methods. It demonstrated improved classification efficiency through comparative evaluation.

Mridula et al., [12] presented an Edge-AI enabled hybrid deep learning framework for intrusion detection in IoT ecosystems. The system achieved accuracy above 98% with low false positive rates. Although focused on botnet intrusion detection, the hybrid architecture is adaptable to fake news detection. The study emphasized lightweight and scalable AI solutions. It demonstrated the effectiveness of edge-based intelligent systems for real-time classification tasks.

III. CHALLENGES

Fake news detection is a complex and evolving problem that involves technical, social, and ethical challenges. Although machine learning and deep learning models have shown high accuracy in controlled experiments, real-world deployment remains difficult. The rapid growth of social media platforms, multilingual content, and AI-generated misinformation increases the complexity of detection systems. Fake news creators continuously modify writing styles, manipulate multimedia content, and exploit user psychology to avoid automated detection. In addition, limited availability of high-quality labeled datasets and issues related to model interpretability further restrict system reliability. Therefore, addressing these challenges is essential for developing robust, scalable, and trustworthy fake news detection frameworks.

1. Data Imbalance and Dataset Quality

One of the major challenges is the imbalance between real and fake news samples in available datasets. Many datasets contain more real news than fake news, which causes models to become biased toward the majority class. Poorly labeled or noisy data also reduces model performance. Additionally, datasets collected from specific domains (such as politics or health) may not generalize well to other topics. High-quality, balanced, and diverse datasets are necessary for improving model accuracy and reliability.

2. Multilingual and Cross-Domain Content

Fake news is generated in multiple languages and across various domains such as politics, health, finance, and entertainment. Most detection models are trained primarily on English datasets, limiting their performance in other languages. Cultural context and regional writing styles further complicate cross-domain adaptation. Developing multilingual and domain-adaptive models remains a significant challenge in global misinformation control.

3. Evolving Writing Styles and Adversarial Attacks

Misinformation creators frequently change their writing patterns to bypass detection systems. They may use sarcasm, coded language, or subtle modifications to avoid keyword-based filters. Adversarial attacks, where small intentional changes are made to fool AI models, can significantly reduce detection accuracy. Continuous model updating and robust learning strategies are required to counter such adaptive threats.

4. Deepfake and Multimedia Manipulation

Modern fake news is not limited to text; it includes manipulated images, videos, and AI-generated deepfake content. Detecting multimedia misinformation requires multimodal learning techniques that combine text, image, and video analysis. However, building such integrated systems demands high computational power and large annotated datasets, making implementation challenging.

5. Real-Time Detection Requirements

Social media platforms generate massive volumes of data every second. Detecting fake news in real time is difficult due to computational constraints and latency issues. Many deep learning models require significant processing time and resources. Designing lightweight and scalable detection frameworks that can operate efficiently in real-time environments is still an open research problem.

6. Lack of Explainability and Transparency

Many advanced detection models, especially deep learning architectures, function as "black boxes." They provide predictions without clear explanations. This lack of interpretability reduces user trust and makes it difficult for policymakers and platform administrators to justify content moderation decisions. Explainable AI techniques are necessary to improve transparency and accountability.

7. Ethical and Freedom of Speech Concerns

Fake news detection systems must balance misinformation control with freedom of expression. Over-filtering may suppress genuine opinions or controversial but factual content. Automated systems can unintentionally introduce bias, affecting certain communities or viewpoints unfairly. Ensuring fairness, neutrality, and ethical compliance is a critical challenge in large-scale deployment.

8. Rapid Spread and Social Influence Patterns

Fake news often spreads faster than real news due to emotional appeal and sensational headlines. Social network structures, user engagement patterns, and bot activities amplify misinformation rapidly. Detecting fake news based solely on textual content may not capture these propagation dynamics. Incorporating social network analysis and user behavior modeling is essential but technically complex.

IV. CONCLUSION

Fake news detection has emerged as a critical research area due to the rapid growth of digital communication and social media platforms. Various machine learning and deep learning techniques, including traditional classifiers and advanced neural network models, have demonstrated high accuracy in identifying misleading content. Graph-based approaches, hybrid architectures, and NLP-driven methods further enhance detection capability by analyzing contextual, structural, and semantic information. However, challenges such as data imbalance, multilingual content, evolving misinformation strategies, deepfake media, and lack of explainability still limit real-world implementation. Future research should focus on developing robust, interpretable, and real-time detection systems that integrate multimodal analysis and ethical AI principles. A balanced combination of technological innovation, policy regulation, and public awareness is essential to effectively combat the global issue of fake news.

REFERENCES

1. S. Mishra, N. Kumar, O. Agarwal, S. Arora and C. Mehta, "Graph Neural Networks for the

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 15, Issue 2, February 2026)

Development of Efficient Fake News Detection," *2025 International Conference on Cognitive Computing in Engineering, Communications, Sciences and Biomedical Health Informatics (IC3ECSBHI)*, Greater Noida, India, 2025, pp. 1011-1015, doi: 10.1109/IC3ECSBHI63591.2025.10990891.

2. T. Ahmad, M. S. Faisal, A. Rizwan, R. Alkanhel, P. W. Khan, and A. Muthanna, "Efficient fake news detection mechanism using enhanced deep learning model," Applied Sciences, vol. 12, no. 3, p. 1743, 2022.
3. Y. M. Rocha, G. A. de Moura, G. A. Desiderio, C. H. de Oliveira, F. D. Lourenco, and L. D. de Figueiredo Nicolete, "The impact of fake news on social media and its influence on health during the covid-19 pandemic: A systematic review," Journal of Public Health, pp. 1–10, 2021.
4. F. Almeida and V. Mortari, "Smartphones, social networks, and fake news: Institutional economics approach to decision making in the twenty-first century," Journal of Economic Issues, vol. 55, no. 2, pp. 342–348, 2021.
5. S. Tasnim, M. M. Hossain, and H. Mazumder, "Impact of rumors and misinformation on covid-19 in social media," Journal of preventive medicine and public health, vol. 53, no. 3, pp. 171–174, 2020.
6. H. T. Phan, N. T. Nguyen, and D. Hwang, "Fake news detection: A survey of graph neural network methods," Applied Soft Computing, p. 110235, 2023.
7. R. Oshikawa, J. Qian, and W. Y. Wang, "A survey on natural language processing for fake news detection," arXiv preprint arXiv: 1811.00770, 2018.
8. S. I. Manzoor, J. Singla, "Fake news detection using machine learning approaches: A systematic review," in 2019 3rd international conference on trends in electronics and informatics (ICOEI). IEEE, 2019, pp. 230–234.
9. J. Li and M. Lei, "A brief survey for fake news detection via deep learning models," Procedia Computer Science, vol. 214, pp. 1339–1344, 2022.
10. Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," IEEE transactions on neural networks and learning systems, vol. 32, no. 1, pp. 4–24, 2020.
11. R. Pandey, P. K. Patidar, P. Verma, G. H. Anjum Khan, S. Harne and R. Tiwari, "A Comparative Study of Random Forest, SVM, and Naive Bayes for Sentiment Analysis Optimization," *2024 IEEE 2nd International Conference on Innovations in High Speed Communication and Signal Processing (IHCSP)*, Bhopal, India, 2024, pp. 1-4, doi: 10.1109/IHCSP63227.2024.10959957.
12. Mridula, S. Shukla, K. Singh, J. Malviya, K. Rawat and R. Tiwari, "Edge-AI Enabled Hybrid Deep Learning Framework for Botnet Intrusion Detection in Modern IoT-Driven Cyber Ecosystems," *2025 5th International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, MANDYA, India, 2025, pp. 1-5, doi: 10.1109/ICERECT65215.2025.11377360.