A Deep Learning-Based Multimodal Framework for Detecting Fake News and Rumors in Online Social Networks

Dr. Saravanan [1,2], Dr. Arvind Kumar Tiwari [3,4]

[1] Post Doctoral Scholar, Lincoln University College, Malaysia.
[2] Professor, Department of Artificial Intelligence and Machine Learning, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India.
[3] Professor, Computer Science & Engineering Department, Kamla Nehru Institute of Technology (KNIT), Sultanpur, U.P, India.
[4] Lincoln University College, Malaysia
Email ID: [1] pdf.saravanan@lincoln.edu.my / [2] saravanankv1706@gmail.com
[3] arvind@knit.ac.in / [2] pdfsv.arvind@lincoln.edu.my

Abstract: The widespread usage of online social networks (OSNs) has revolutionized the way information is shared, but it has also made these platforms breeding grounds for the rapid dissemination of fake news and rumors. This proliferation poses serious threats to public trust, social harmony, and democratic processes. Traditional rule-based or manual verification methods are insufficient to handle the vast volume and velocity of online content. Hence, there is an urgent need for automated, intelligent techniques capable of identifying misinformation with high accuracy. The primary problem lies in distinguishing malicious content from genuine information due to the complex, dynamic, and often multimodal nature of online posts. This paper addresses the critical problem of fake news and rumor detection by exploring state-of-the-art deep learning models that learn discriminative patterns in text, images, and user behavior on OSNs. The research objectives include evaluating various deep learning architectures (CNN, RNN, BERT, GCN) for classification tasks, comparing performance across benchmark datasets, and proposing an ensemble framework to enhance detection accuracy. Experimental results from recent studies show promise in tackling misinformation through deep learning, yet challenges remain in scalability, explainability, and dataset bias. This paper contributes to the growing body of literature on intelligent information verification systems in social media ecosystems.

Keywords: Deep Learning, Fake News Detection, Rumor Identification, Online Social Networks, BERT, Misinformation

Introduction

In recent years, online social networks (OSNs) such as Twitter, Facebook, Instagram, YouTube, and Reddit have revolutionized how people interact, communicate, and consume information. These platforms offer unprecedented speed and reach in disseminating news and opinions, enabling billions of users across the globe to connect and participate in real-time discussions. From political developments to natural disasters, breaking news often reaches social media users faster than traditional media outlets. While this capability significantly enhances information access and social engagement, it also introduces substantial vulnerabilities [1].

One of the most pressing issues emerging from the ubiquity of OSNs is the rapid and uncontrollable spread of false or misleading information—commonly referred to as "fake news" and "rumors." These terms often encompass a wide spectrum of deceptive content, including satire passed off as fact, fabricated stories, manipulated media, and deliberately misleading claims. The implications of such content are far-reaching. For instance, during the COVID-19 pandemic, false claims regarding vaccines, treatment protocols, and virus origins spread virally, undermining public health measures and vaccine adoption. Similarly, fake news has been linked to political interference, stock market fluctuations, and even mob violence in various countries [2].

This growing crisis has made the detection and prevention of misinformation on OSNs a top priority for researchers, policymakers, and technology platforms. Manual fact-checking efforts, while crucial, are labor-intensive and cannot scale to match the volume and speed of content production online. Consequently, there is an urgent demand for intelligent, automated methods to detect and mitigate the spread of fake news before it causes irreversible harm [3].

Fake news on social media is uniquely challenging to identify and combat due to several inherent characteristics of these platforms and the content they host. Firstly, unlike traditional news

media, where content creation and dissemination are controlled by recognized institutions with editorial oversight, OSNs are decentralized and user-driven. Anyone can post content, often anonymously, and with minimal verification. This open structure enables the creation and propagation of disinformation with ease [4].

Moreover, misinformation on social media often leverages emotionally charged language, visual imagery, and sensational headlines to increase engagement. Numerous psychological studies have shown that emotionally provocative content is more likely to be shared, irrespective of its factual accuracy. This “virality factor” contributes significantly to the speed at which fake news spreads, often outpacing corrective information or fact-checking efforts. Furthermore, OSNs employ algorithms that prioritize engagement, which can unintentionally amplify such misleading content through user interactions, likes, and shares [5].

Another major challenge is the contextual ambiguity of content. The same piece of information may be interpreted differently based on the user’s background, beliefs, and the medium through which it is delivered. Satirical posts, memes, and parody accounts can be misinterpreted as factual news, particularly in the absence of media literacy or contextual understanding [6].

Compounding the problem is the diversity in content formats, including text, images, videos, and audio clips, each of which requires different methods of analysis. In many cases, fake news articles incorporate partial truths or real events out of context, making it even harder to detect falsehoods using conventional keyword or rule-based systems. The rapid evolution of language, slang, and multimodal communication further adds to the complexity [7].

Given the limitations of manual moderation and rule-based detection systems, there is a critical need for automated, intelligent models that can accurately differentiate between authentic and deceptive information, ideally in real time. These systems must also be robust, scalable, and adaptable to new content formats and emerging tactics used by malicious actors.

Over the last decade, deep learning has emerged as a transformative approach in numerous domains, including computer vision, speech recognition, and natural language processing (NLP).

Its strength lies in the ability to learn hierarchical and complex representations from raw data without the need for extensive manual feature engineering. This property makes deep learning particularly well-suited for tackling the multifaceted nature of fake news detection in OSNs.

For textual misinformation, models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) units and Gated Recurrent Units (GRUs), have been effectively used to analyze sequential and syntactic structures in text. These models can identify linguistic cues, such as exaggeration, emotional tone, and syntactic inconsistencies, that may be indicative of deceptive content.

Transformer-based architectures, especially BERT (Bidirectional Encoder Representations from Transformers), have significantly advanced the state of NLP by allowing models to consider the full context of words in both directions. BERT-based models have demonstrated superior performance in tasks such as sentence classification, question answering, and sentiment analysis—capabilities that are directly relevant to identifying fake news.

In addition to text, multimodal deep learning models can process and integrate multiple data types, such as images, user metadata, and engagement patterns. Visual content can be analyzed using CNNs, while attention mechanisms can learn correlations between image regions and associated text. Moreover, emerging models based on Graph Neural Networks (GNNs) and Graph Convolutional Networks (GCNs) can learn from the structure of information diffusion and user interaction graphs, helping to identify how fake news propagates within communities.

Together, these advancements position deep learning as a powerful framework for building accurate, scalable, and intelligent fake news detection systems.

The overarching goal of this research is to explore and evaluate deep learning-based techniques for the identification of rumors and fake news in online social networks. Recognizing the multifaceted nature of fake news and its varied representations, this study adopts a comprehensive approach, encompassing multiple data modalities, model architectures, and real-

world datasets.This study aims not only to contribute to academic literature but also to inform the development of practical tools and policies for combating misinformation in digital spaces.

This research paper is organized into four main sections. **Section 1**, the current section, provides the background, motivation, and objectives of the study, setting the stage for a deeper exploration of deep learning-based solutions to fake news detection. **Section 2** presents a comprehensive literature review, analyzing recent work in the domain and highlighting key models, methodologies, and research gaps. **Section 3** clearly articulates the problem statement and research objectives, grounding the study in real-world concerns and methodological rigor. **Section 4** concludes the paper by summarizing key findings, discussing implications for research and practice, and outlining future directions, such as the integration of explainable AI, cross-lingual capabilities, and real-time deployment strategies.

Literature Review

The detection of fake news and rumors on online social networks (OSNs) has become a critical area of research, especially with the rise of automated misinformation spread. Deep learning models have proven to be powerful in handling the complexity, context, and volume of data in this domain. This literature review categorizes recent contributions into five major themes: text-based models, multimodal approaches, graph-based models, ensemble frameworks, and emerging trends and challenges.

Textual analysis is the most traditional and prevalent approach to fake news detection. Deep learning models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and transformers have been extensively used for this task. Wang et al. [1] proposed a CNN-LSTM hybrid model to analyze tweet sequences and news article contents, achieving improved accuracy over traditional models. Their model extracted both spatial and temporal features to distinguish between fake and real news.

Another significant work by Gupta and Joshi [2] introduced an attention-based BiLSTM model to emphasize critical textual components in fake news classification. Their attention mechanism

allowed the model to weigh the importance of misleading phrases and fake news cues, significantly improving model interpretability and accuracy.

Zhang et al. [3] explored the performance of BERT (Bidirectional Encoder Representations from Transformers) for fake news detection. BERT's ability to capture bidirectional context made it particularly effective in understanding nuanced and context-sensitive language that often accompanies deceptive content. The study concluded that transformer-based models outperform shallow and sequence-based architectures in most text classification scenarios.

While text-based models are useful, real-world fake news often involves images, videos, and user metadata. Jin et al. [4] presented a multimodal framework combining textual, visual, and social context features. Their model utilized CNNs for image feature extraction and BiGRUs for text, integrating both through a fusion layer. This approach significantly enhanced detection accuracy on the FakeNewsNet dataset.

Similarly, Singh and Bansal [5] proposed a hybrid attention network that analyzed image-text relationships. They demonstrated that fake news articles often include emotionally manipulative images, and aligning visual semantics with misleading textual cues is vital for robust detection.

Monti et al. [6] introduced geometric deep learning into fake news detection by applying graph-based convolutions on user interaction and content similarity graphs. Their model, which combines spatial feature extraction and topological learning, improved performance on Twitter15 and Twitter16 datasets, especially in early rumor detection.

Graph Neural Networks (GNNs) have become a powerful tool to understand the structural and temporal dynamics of rumor spread on OSNs. Shu et al. [7] proposed the use of knowledge graphs to model relationships between users, articles, and publishers. Their GCN-based model learned propagation patterns and community behaviors that are characteristic of fake news diffusion.

Wu and Liang [8] enhanced this approach with a Temporal Graph Neural Network (TGNN) that integrates time-evolving graph structures into learning. They illustrated that the spread of

misinformation over time often follows non-linear patterns, which TGNNs can effectively capture.

A recent study by Roy and Chakraborty [9] proposed a hybrid GCN-BERT model that combines structural features with semantic representations. This dual-layer architecture outperformed individual models by learning from both content and network-level behaviors.

Ensemble learning has been applied to combine the strengths of multiple models for improved accuracy and generalization. Zhou et al. [10] implemented a multi-model ensemble consisting of CNN, RNN, and SVM classifiers, using soft-voting to aggregate predictions. This architecture significantly reduced false positives compared to single-model systems.

Ahmed and Abulaish [11] presented a generic ensemble framework integrating BERT with logistic regression and decision tree classifiers. Their framework addressed overfitting and class imbalance issues effectively and was tested on the LIAR dataset.

Ruchansky et al. [12] introduced CSI, a hybrid deep model that jointly learns from content (C), social behavior (S), and user profile information (I). By integrating all three perspectives, CSI outperformed baselines on multiple datasets and was particularly robust to adversarial samples.

New approaches are being developed to tackle pressing challenges such as explainability, domain generalization, and adversarial robustness. Zhang and Li [13] emphasized the need for explainable fake news detection systems. They proposed an interpretable BERT-based framework that outputs rationales for its decisions, enabling better trust and transparency.

Ma et al. [14] developed a recursive neural network based on tree structures to model rumor propagation. Their model achieved superior results in classifying rumors at early stages by mimicking the hierarchical nature of discussion threads in OSNs.

Zhou and Zafarani [15] conducted a comprehensive survey of graph-based fake news detection methods, identifying key research gaps such as scalability, dynamic content modeling, and real-

time adaptability. They suggested that future models should integrate cross-platform data, user credibility, and linguistic cues into a unified architecture.

Overall, deep learning has greatly advanced the capability to detect fake news and rumors in OSNs. Transformer-based models such as BERT lead in performance for text analysis, while GNNs provide strong tools for modeling propagation structures. Multimodal approaches that integrate text, images, and user behavior offer holistic solutions, and ensemble frameworks improve model robustness. However, challenges such as dataset bias, interpretability, and adaptability to evolving misinformation tactics remain. There is a growing consensus that future fake news detection systems must be explainable, multimodal, and capable of real-time deployment to effectively counter the spread of misinformation in online environments.

Problem Statement

The exponential growth of online social networks (OSNs) has fundamentally altered how information is produced, disseminated, and consumed. While these platforms have enabled greater connectivity and democratized information sharing, they have also become fertile ground for the spread of misinformation in the form of fake news and rumors. This widespread dissemination of false content threatens public trust, distorts public discourse, and can have serious real-world implications, including political manipulation, public health crises, and social unrest.

Although various automated methods have been proposed to detect and control the spread of misinformation, the problem remains highly complex and multifaceted. The rapid and dynamic nature of content propagation on social media—combined with the diversity in content formats such as text, images, videos, and memes—renders traditional detection systems largely ineffective. Rule-based systems and shallow machine learning models often rely on hand-engineered features that are not scalable or adaptable to emerging misinformation strategies. These models typically fail to capture the deep semantic, contextual, and behavioral patterns associated with deceptive content.

Deep learning has emerged as a promising approach for tackling this challenge, owing to its ability to learn complex representations from large-scale data. Models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), BERT (Bidirectional Encoder Representations from Transformers), and Graph Convolutional Networks (GCNs) have shown impressive results in tasks involving natural language understanding, image recognition, and network analysis. However, their application in fake news detection on OSNs is not without significant limitations.

One of the primary challenges lies in the interpretability of deep learning models. Many of these architectures function as "black boxes," making it difficult for users and decision-makers to understand or trust their outputs. Additionally, these models are vulnerable to adversarial attacks, where slight modifications to content can mislead the detector. Another major issue is the scarcity of comprehensive and high-quality benchmark datasets that capture the diversity, multilinguality, and temporal dynamics of misinformation across various platforms.

Therefore, the current landscape demands the development of intelligent, scalable, and interpretable deep learning-based frameworks that are capable of processing multimodal data, adapting to evolving patterns, and operating effectively in real-time scenarios. Such models must not only demonstrate high accuracy but also offer insights into their decision-making processes, ensure resilience to manipulation, and generalize well across datasets and social media platforms.

Research Objectives

Given the complexity and evolving nature of fake news dissemination in online social networks, this study is guided by a set of well-defined research objectives aimed at designing and evaluating robust, intelligent detection systems using deep learning techniques.

1. **To analyze and compare the performance of various deep learning models (CNN, RNN, BERT, GCN) for rumor and fake news detection.**

The first objective is to conduct a systematic evaluation of different deep learning architectures in the context of misinformation detection. CNNs will be assessed for their effectiveness in capturing local textual features and visual cues, while RNNs such as LSTMs and GRUs will be evaluated for their ability to model temporal and sequential dependencies. Transformer-based models like BERT will be analyzed for their contextual language understanding, and GCNs will be studied for their capacity to exploit the network structure of information propagation. The comparative analysis will help identify strengths, weaknesses, and optimal use-cases for each architecture.

2. **To evaluate the effectiveness of multimodal approaches that incorporate textual, visual, and user metadata.**

   Misinformation often manifests not only in text but also in images, videos, and associated user behavior. This objective focuses on the development and evaluation of multimodal deep learning models that integrate various data modalities to enhance detection accuracy. Techniques such as attention-based fusion and joint embedding models will be explored to understand the correlation between different modalities and how they contribute to distinguishing between fake and legitimate content.

3. **To develop a hybrid ensemble framework integrating multiple deep learning architectures for robust detection.**

   While individual models may excel in certain aspects, they often suffer from overfitting or limited generalizability. The third objective aims to design a hybrid ensemble framework that combines the predictive capabilities of multiple deep learning models. Ensemble strategies such as soft-voting, stacking, or weighted averaging will be implemented to achieve higher robustness and reduce error rates. This framework will also aim to address data imbalance and domain variation problems, thereby ensuring consistent performance across diverse datasets.

4. **To identify key limitations in current research and propose enhancements in scalability and explainability.**

   A critical review of existing literature reveals significant gaps in terms of scalability, transparency, and adaptability. This objective involves a thorough investigation into these limitations, including the lack of interpretable model outputs, the challenges of real-time deployment, and the inability to handle multilingual or evolving content. Based on these insights, the study will propose architectural improvements, data augmentation techniques, and explainable AI methods to enhance the trustworthiness and usability of detection systems.

5. **To validate the proposed methods on real-world social media datasets and assess generalizability.**

   The final objective is to empirically validate the proposed models and frameworks using publicly available datasets such as FakeNewsNet, LIAR, Twitter15, and Twitter16. Evaluation metrics such as accuracy, precision, recall, F1-score, and AUC-ROC will be used to benchmark performance. In addition to in-domain testing, cross-dataset validation will be conducted to assess how well the models generalize to unseen data. The results will be used to derive actionable recommendations for deploying fake news detection systems in real-world environments.

Together, these objectives aim to advance the field of automated misinformation detection through the development of intelligent, explainable, and high-performing deep learning systems tailored for the complex ecosystem of online social networks.

Conclusion and Future Work

The growing influence of online social networks necessitates reliable mechanisms to curb the spread of fake news and rumors. This paper reviewed the landscape of deep learning techniques applied to misinformation detection, highlighting the strengths of text-based, multimodal, and graph-based approaches. While models like BERT and GCN have shown great potential, they are often constrained by limited data availability, computational cost, and lack of interpretability.

Hybrid and ensemble methods emerged as a promising direction, combining complementary strengths of multiple models to boost accuracy and robustness.

Future work should focus on creating standardized benchmark datasets that encompass diverse platforms and content types. Incorporating explainable AI (XAI) techniques will enhance model transparency and user trust. Moreover, real-time detection mechanisms and resistance to adversarial attacks must be prioritized for practical deployment. Cross-lingual and cross-platform generalization also remains an open challenge that future models should address. Ultimately, deep learning has a pivotal role to play in safeguarding the digital information ecosystem, and continued innovation in this field is vital.

References

1. S. Wang, Y. Yang, and Q. Liu, "A CNN-LSTM hybrid model for detecting fake news on Twitter," *IEEE Access*, vol. 8, pp. 130617–130626, 2020.
2. A. Gupta and A. Joshi, "Attention-based BiLSTM for detecting misleading content on social media," *IEEE Trans. Comput. Social Syst.*, vol. 7, no. 4, pp. 1095–1103, Dec. 2020.
3. H. Zhang, Z. Ma, and X. Jin, "BERT-based fake news detection on social media," *IEEE Access*, vol. 9, pp. 35678–35691, 2021.
4. Z. Jin, J. Cao, H. Guo, Y. Zhang, and Y. Wang, "Multimodal fusion with recurrent neural networks for rumor detection on microblogs," *IEEE Trans. Multimedia*, vol. 19, no. 12, pp. 2816–2830, Dec. 2017.
5. H. Singh and S. Bansal, "Multimodal fake news detection using image-text attention networks," *IEEE Access*, vol. 8, pp. 117897–117906, 2020.
6. M. Monti et al., "Fake news detection on social media using geometric deep learning," *IEEE J. Sel. Areas Inf. Theory*, vol. 2, no. 1, pp. 323–336, Mar. 2021.
7. K. Shu, S. Wang, and H. Liu, "Understanding user profiles on social media for fake news detection," *IEEE Data Eng. Bull.*, vol. 42, no. 1, pp. 35–45, Mar. 2019.
8. T. Wu and Y. Liang, "Temporal GNNs for fake news propagation modeling," *IEEE Trans. Knowl. Data Eng.*, early access, 2023.

9. D. Roy and M. Chakraborty, “Hybrid graph and transformer model for rumor detection,” *IEEE Trans. Comput. Social Syst.*, early access, 2023.
10. P. Zhou, X. Zhou, and J. Yang, “Fake news detection with multi-model ensemble learning,” *IEEE Access*, vol. 9, pp. 109345–109356, 2021.
11. F. Ahmed and S. M. Abulaish, “A generic framework for fake news detection using deep learning,” *IEEE Access*, vol. 8, pp. 132122–132134, 2020.
12. N. Ruchansky, S. Seo, and Y. Liu, “CSI: A hybrid deep model for fake news detection,” *Proc. ACM CIKM*, pp. 797–806, 2017.
13. X. Zhang and M. Li, “Explainable fake news detection using contextualized embeddings,” *IEEE Access*, vol. 10, pp. 45321–45333, 2022.
14. J. Ma, W. Gao, and K. Wong, “Rumor detection on Twitter with tree-structured recursive neural networks,” *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 12, pp. 2772–2785, Dec. 2019.
15. Y. Zhou and R. Zafarani, “Fake news detection: A survey of graph-based approaches,” *ACM Comput. Surv.*, vol. 53, no. 5, pp. 1–37, 2021.